

# EVALUATION OF HUMAN WORK

FOURTH EDITION

Edited by  
**JOHN R. WILSON**  
**SARAH SHARPLES**





EVALUATION  
OF  
HUMAN WORK  
FOURTH EDITION





# EVALUATION OF HUMAN WORK

FOURTH EDITION

Edited by  
JOHN R. WILSON  
SARAH SHARPLES



CRC Press

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

CRC Press  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC  
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works  
Version Date: 20150309

International Standard Book Number-13: 978-1-4665-5962-2 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Visit the Taylor & Francis Web site at**  
**<http://www.taylorandfrancis.com>**

**and the CRC Press Web site at**  
**<http://www.crcpress.com>**

*To John. We all miss you and thank you for your friendship  
and contribution to our work and lives.*

*To Craig, Ellen and Luke. Without you, none of this would have been possible.*

xxx



---

# Contents

|                                              |       |
|----------------------------------------------|-------|
| Preface to the Fourth Edition.....           | xi    |
| Acknowledgements to the Fourth Edition ..... | xiii  |
| Preface to the Third Edition .....           | xv    |
| Acknowledgements to the Third Edition .....  | xvii  |
| Preface to the Second Edition.....           | xix   |
| Acknowledgements to the Second Edition ..... | xxi   |
| Preface to the First Edition .....           | xxiii |
| Acknowledgements to the First Edition.....   | xxv   |
| Editors .....                                | xxvii |
| Contributors .....                           | xxix  |

|                                                                      |   |
|----------------------------------------------------------------------|---|
| <b>Chapter 1</b> Methods in the Understanding of Human Factors ..... | 1 |
| <i>John R. Wilson and Sarah Sharples</i>                             |   |

## ***SECTION I Fundamental Approaches and Tools***

|                                                                                  |     |
|----------------------------------------------------------------------------------|-----|
| <b>Chapter 2</b> Experimental Design and Analysis.....                           | 37  |
| <i>Colin Drury, Victor Paquet and Harrison Kelly</i>                             |     |
| <b>Chapter 3</b> Study and Analysis of Complex Cognitive Work.....               | 61  |
| <i>Ann Bisantz, Emilie M. Roth and Jennifer Watts-Englert</i>                    |     |
| <b>Chapter 4</b> Methods for Collecting and Observing Participant Responses..... | 83  |
| <i>Sarah Sharples and Sue Cobb</i>                                               |     |
| <b>Chapter 5</b> Qualitative Methodology.....                                    | 119 |
| <i>Sue Hignett and Hilary McDermott</i>                                          |     |
| <b>Chapter 6</b> Task Analysis .....                                             | 139 |
| <i>Andrew Shepherd</i>                                                           |     |
| <b>Chapter 7</b> Knowledge Elicitation: Methods, Tools and Techniques .....      | 163 |
| <i>Nigel R. Shadbolt and Paul R. Smart</i>                                       |     |
| <b>Chapter 8</b> Simulation and Digital Human Modelling.....                     | 201 |
| <i>Glyn Lawson and Gary Burnett</i>                                              |     |

## **SECTION II    *Methods for Design of Work***

|                   |                                                                                     |     |
|-------------------|-------------------------------------------------------------------------------------|-----|
| <b>Chapter 9</b>  | Systems Analysis and Design.....                                                    | 221 |
|                   | <i>Robert Houghton, Nora Balfe and John R. Wilson</i>                               |     |
| <b>Chapter 10</b> | Involving People in Design Research.....                                            | 249 |
|                   | <i>David Kirk, Ian McClelland and Jane Fulton Suri</i>                              |     |
| <b>Chapter 11</b> | Inclusive Design and Design for Special Populations.....                            | 299 |
|                   | <i>Edward Elton and Colette Nicolle</i>                                             |     |
| <b>Chapter 12</b> | Analysis and Design of Warnings in the Workplace.....                               | 331 |
|                   | <i>Christopher B. Mayhorn, Michael S. Wogalter and Kenneth R. Laughery</i>          |     |
| <b>Chapter 13</b> | Evaluating Human-Computer Interaction.....                                          | 359 |
|                   | <i>Chris Baber</i>                                                                  |     |
| <b>Chapter 14</b> | Analysing and Designing Control Facilities .....                                    | 383 |
|                   | <i>Rebecca Charles, Sarah Sharples, Jane A. Rajan, John R. Wilson and John Wood</i> |     |

## **SECTION III    *Assessment of Impact of Work on Individuals***

|                   |                                                    |     |
|-------------------|----------------------------------------------------|-----|
| <b>Chapter 15</b> | Physiological Bases of Work Assessment .....       | 419 |
|                   | <i>Karen Sjøgaard and Gisela Sjøgaard</i>          |     |
| <b>Chapter 16</b> | Assessment of Physical Job Demands.....            | 447 |
|                   | <i>Jonisha P. Pollard and Patrick G. Dempsey</i>   |     |
| <b>Chapter 17</b> | Biomechanical Basis of Ergonomics Assessment.....  | 487 |
|                   | <i>Patrick G. Dempsey and Moira Munro</i>          |     |
| <b>Chapter 18</b> | Definition and Measurement of Human Workload ..... | 515 |
|                   | <i>Sarah Sharples and Ted Megaw</i>                |     |
| <b>Chapter 19</b> | Situation Awareness .....                          | 549 |
|                   | <i>David Golightly</i>                             |     |

|                   |                                                                                                     |     |
|-------------------|-----------------------------------------------------------------------------------------------------|-----|
| <b>Chapter 20</b> | Assessment of Potentially Stressful Work and Organisational Environments:<br>Theory and Method..... | 565 |
|                   | <i>Tom Cox and Amanda Griffiths</i>                                                                 |     |
| <b>Chapter 21</b> | Psychophysiological Measurement of Physical and Cognitive Work.....                                 | 583 |
|                   | <i>Jung-Yong Kim, Young-Jin Cho and Seung-Nam Min</i>                                               |     |
| <b>Chapter 22</b> | Neuroergonomic Methods for the Evaluation of Physical and Cognitive Work .....                      | 609 |
|                   | <i>Raja Parasuraman and Ranjana Mehta</i>                                                           |     |

## ***SECTION IV    Assessment and Design of the Physical Workplace***

|                   |                                                        |     |
|-------------------|--------------------------------------------------------|-----|
| <b>Chapter 23</b> | The Environmental Ergonomics Survey .....              | 641 |
|                   | <i>Ken Parsons</i>                                     |     |
| <b>Chapter 24</b> | Ergonomics Assessment of Thermal Environments .....    | 655 |
|                   | <i>Ken Parsons</i>                                     |     |
| <b>Chapter 25</b> | Assessment of the Visual Environment .....             | 677 |
|                   | <i>Peter A. Howarth</i>                                |     |
| <b>Chapter 26</b> | Auditory Environment and Noise Assessment.....         | 705 |
|                   | <i>Christine M. Haslegrave</i>                         |     |
| <b>Chapter 27</b> | Anthropometry for Ergonomic Design .....               | 725 |
|                   | <i>David Feathers, Clive D'Souza and Victor Paquet</i> |     |

## ***SECTION V    Methods in Organisational and Social Human Factors***

|                   |                                                          |     |
|-------------------|----------------------------------------------------------|-----|
| <b>Chapter 28</b> | Sociotechnical Design of Work Systems.....               | 753 |
|                   | <i>Patrick Waterson</i>                                  |     |
| <b>Chapter 29</b> | Teamwork and Team Performance Measurement.....           | 773 |
|                   | <i>Alyssa Marshall, Tiffany Bisbey and Eduardo Salas</i> |     |

|                   |                                                                              |     |
|-------------------|------------------------------------------------------------------------------|-----|
| <b>Chapter 30</b> | Human Reliability Assessment .....                                           | 791 |
|                   | <i>Barry Kirwan</i>                                                          |     |
| <b>Chapter 31</b> | Incident Reporting and Analysis .....                                        | 821 |
|                   | <i>Brendan Ryan</i>                                                          |     |
| <b>Chapter 32</b> | Ergonomic Interventions in the Implementation of New Technical Systems ..... | 837 |
|                   | <i>Ken Eason</i>                                                             |     |
| <b>Chapter 33</b> | Methods in Systems Ergonomics .....                                          | 855 |
|                   | <i>Carys E. Siemieniuch and Murray A. Sinclair</i>                           |     |

## **SECTION VI    *Implementation and Analysis of Human Factors***

|                   |                                                     |     |
|-------------------|-----------------------------------------------------|-----|
| <b>Chapter 34</b> | Participatory Ergonomics .....                      | 883 |
|                   | <i>Diane Gyi, Sally Shalloe and John R. Wilson</i>  |     |
| <b>Chapter 35</b> | Cost–Benefit Analysis for Human Factors .....       | 907 |
|                   | <i>Emile Tompa, Mike Foley and Dwayne Van Eerd</i>  |     |
| <b>Chapter 36</b> | Integrating Ergonomics and Quality Concepts .....   | 931 |
|                   | <i>Jörgen A.E. Eklund and Paul H.P. Yeow</i>        |     |
| <b>Chapter 37</b> | Standards in Ergonomics Design and Evaluation ..... | 957 |
|                   | <i>Mike Gray and Tom Stewart</i>                    |     |

## **SECTION VII    *Conclusion***

|                   |                                                           |     |
|-------------------|-----------------------------------------------------------|-----|
| <b>Chapter 38</b> | Ergonomics and Human Factors as Reflective Practice ..... | 975 |
|                   | <i>Sarah Sharples and Peter Buckle</i>                    |     |



---

# Preface to the Fourth Edition

It was an honour and a privilege to be asked by John to join him as an editor of the fourth edition of this text, and it is with great sadness that I have completed the task alone since John's death in July 2013. Before his death, we had discussed the revised content and structure of the book, and John had led the commissioning of the revised chapters to be submitted. He had also begun work on revising many of the chapters on which he was a co-author, and we had discussed some of the changes we would make to the structure and content of the text. We particularly discussed the introductory Chapter 1, which provides an overview of our approach to methods in Ergonomics/Human Factors (E/HF).

In reviewing the previous editions of this book (and it is worth noting that the first edition was published whilst I was still at school and I am ashamed to say I had not yet heard of ergonomics), it is clear to me that this edition is in fact the most comprehensive update that the book has had. All chapters have been revised, most of them with significant changes of authorship or content. It is interesting to see many other authors, perhaps like me, who have worked with some of the original authors of the chapters over recent years and are now contributing to the writing of this text. There are five completely new chapters, covering inclusive design (Chapter 11), situation awareness (Chapter 19), neuroergonomics (Chapter 22), ergonomics and quality (Chapter 36) and standards in ergonomics design and evaluation (Chapter 37). The final chapter (Chapter 38) is also new and considers the notion of reflective practice, which underpins our work as E/HF practitioners. Most chapters have new authors and co-authors, and much of the content is substantially changed and updated. Three chapters on which John was originally the lead author (Chapters 9, 14 and 35) have been taken on by new teams.

Much continues to change in the discipline of E/HF. John highlighted the interest in the systems perspective in the preface to the third edition of this text; this systems perspective is now prevalent within much E/HF practice and underpins many of the discussions within this book, as well as being the particular focus of Chapters 9, 28 and 33. Further trends include the increased availability and utility of physiological devices. Tools that were once solely for use in a laboratory context are now feasible to be applied in field contexts; it is therefore essential that we understand the nature of these tools, the quality of output they deliver and the relationship between these physiologically measured parameters and the theoretical concepts, such as workload, fatigue, stress or effort, to which we wish to relate them.

The name of this text is, and has been for the three previous editions, *Evaluation of Human Work*. In the preface to the first edition, the editors note that 'we are primarily concerned with people at work' but that this does not rule out 'contributions relevant to people's activities at home, leisure or on the road'. This is even more important in today's society with its blurring of the home–work boundary – not only do we now see many people who work from a home environment, but we also see how work technologies are appropriated for home use, and vice versa.

A final extremely important point to note, which reflects the increased and recognised importance of the discipline of E/HF, is the 2014 award of a Royal Charter to the United Kingdom's Institute of Ergonomics and Human Factors. This status is only awarded to those organisations that represent a field that is 'unique' and represent a tremendous endorsement of the work that E/HF practitioners do all over the world to improve the way in which people work and interact with systems and technologies.

I read the first edition of the book as a student of human factors in 1994, little expecting to one day be its editor. Its appeal to me was in its detail, opinion and guidance – at the end of reading a chapter, I felt as though I could understand the context of 'how' to practise ergonomics and what

aspects of method selection and implementation I should consider. Whilst working on the book I have been both intimidated and pleased by the response of those who have commented on it being their favourite ergonomics/human factors text. I hope that this revision meets their expectations and retains the unique place the text has amongst the rich set of books that are now available for those practising E/HF.

Some of the elements of the book were very much John's and reflected the thoughts that I am sure he and Nigel Corlett, the co-editor of the first three editions, discussed and debated many times. I was fortunate to also engage in such debates with John, and there were certainly points on which we did not always agree! I have tried to reflect his views where I can but also have included views and interpretations of my own; occasionally, and particularly in Chapter 1, I note this explicitly in footnotes, sometimes interchanging between 'I' and 'we' as pronouns.

This book originally emerged from a now infamous gathering of some of the leading minds of the world of ergonomics for the Second International Occupational Ergonomics Symposium in the mid-1980s at Zadar, Yugoslavia. Just a few weeks before the submission of this volume, it was lovely to gather many other like-minded experts in E/HF, including some of those who attended that original meeting, for a conference on human factors in complex systems, organised in memory of John, at the University of Nottingham. Along with recollections of John and his contribution to E/HF, there was much lively critique and debate of many of the concepts presented in this text; some of these are reflected in the final chapter of this book.

**Sarah Sharples**

MATLAB® is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc.  
3 Apple Hill Drive  
Natick, MA 01760-2098 USA  
Tel: 508-647-7000  
Fax: 508-647-7001  
E-mail: [info@mathworks.com](mailto:info@mathworks.com)  
Web: [www.mathworks.com](http://www.mathworks.com)

---

# Acknowledgements to the Fourth Edition

It has been a difficult but immensely rewarding task to complete this book, and it would not have been possible without the support of all of the chapter authors, who have been so diligent in updating and writing their chapters and responding to my requests throughout the editorial process. In particular, Colin Drury, who has provided support and content for this volume from its first edition, has been a tremendous moral supporter and advisor to me in the final months leading up to submission of the text. The publishers at Taylor & Francis Group/CRC Press, especially Cindy Carelli, Paul Abraham and Jennifer Ahringer, have been very patient with responding to my requests and supporting me through the mammoth task of collating and submitting this volume to them.

I have had significant support from the University of Nottingham in enabling me to squeeze the time to complete this book. I particularly thank the Department of Mechanical, Materials and Manufacturing Engineering for allowing me to cover some of my teaching duties whilst working on the book and thank those who delivered my teaching and administration duties in my partial absence. The research grants EP/G065802/1, EP/G037574/1 and EP/K014161/1 provided financial support to enable this work.

I am particularly thankful to colleagues in the Human Factors Research Group for their encouragement and work to help me complete this book that has been so important and useful to so many of us over the years. Many group members helped review the content of the methods table in Chapter 1 (in particular, Sally Shalloe, Alyson Langley, Harshada Patel, Julie Waldron, Brendan Ryan, Laura Lewis and Rob Houghton provided useful input and critique). Thanks also to Cath Harvey, Michael Brown, Genovefa Kefalidou, Liz Dowthwaite, Lizzie Crundall, Anne Floyd, Richard Eastgate and Will Knight for help with proofreading. Thank you also to Lynne Mills who provided me with some aspects of the ‘history’ of the volume which were very useful when putting all of the elements together. Credit must also go to Heidi Freestone who has battled with my diary and protected my time as much as she could, dealing very patiently with the different demands she received, keeping my students calm and responding to my numerous last-minute requests!

I have also received informal support from many colleagues within the E/HF community, particularly Chris Baber and Peter Buckle, who have provided me with informal comments and thoughts about the book that have proved very useful to my editorial process. Thanks also to David Gilmore who, perhaps unwittingly, over a glass of wine at the annual Ergonomics and Human Factors conference in the United Kingdom in 2014, provided us with the last line of the final chapter of this book.

I had worked with John for almost 20 years (which proved helpful when having to interpret his handwritten notes on drafts of the chapters!), and this is still very much his book. His legacy to the discipline of ergonomics and human factors is primarily through his influence on the people he taught and worked with but is also through his writing. This text is probably the most important of all of the things he wrote, and I hope he would have been proud of this fourth edition.

Finally, I thank my husband, Craig, and my children, Ellen and Luke, for their love and support during what has been a very busy and demanding time. You have been extremely patient whilst I have worked in the evenings and weekends. I’m sure you will be very pleased to hear that the book is finally finished!

**Sarah Sharples**



---

# Preface to the Third Edition

Ergonomics/human factors is a fast-moving discipline. The domains in which we work and can see demonstrable success from the application of our knowledge and skills are constantly expanding. At the same time we are continually improving and enlarging our methodological base, developing new theories, approaches, methods and tools as well as refining ones we have used before. This third edition of *Evaluation of Human Work* has been published now to try to reflect many of these changes. Therefore in every case bar one, the chapters from the second edition have been substantially revised, and there are several new chapters.

In general, the completely new chapters reflect the growth of interest in a systems viewpoint within ergonomics/human factors, and particularly understanding physical and cognitive work in the context of the social and organisational setting in which it takes place. Thus the new chapters include ones on sociotechnical design of work systems (Chapter 29), team design and evaluation (Chapter 30), learning from failures through a joint cognitive systems perspective (Chapter 34), and the analysis of organisational processes (Chapter 38).

In addition to these new chapters, several chapters have not just been revised in terms of updating material but with a completely new approach. Thus the chapter concerned with user trials in the second edition has been expanded into one which addresses many techniques in user-centred design (Chapter 11). The chapter on knowledge elicitation for expert systems has been rewritten to reflect the increased interest in understanding the nature of knowledge and of knowledge management in contemporary systems (Chapter 8). The opening chapters on assessment and design of the physical workplace to do with the environmental factors of climate, visual conditions and noise have been accompanied by a short chapter discussing environment surveys generally (Chapter 22). Finally, reflecting what is current practice, the chapter on accident reporting and analysis has been replaced by one on systems for near miss reporting and analysis (Chapter 33).

One thing which has remained unchanged from the first and second editions is that this text is produced *not* as a cookbook of ergonomics methods. Whilst there is a place for handbooks and manuals which describe each of hundreds of different ergonomics/human factors tools and techniques, with some form of lookup table for users to select amongst them, that is not the intention and thrust of this book. Rather the intention is to place ergonomics methodology in context, and each chapter carefully describes the background to method development in that area and to the application of methods and tools. In this way, it is intended to make the text suitable for teaching on ergonomics and human factors courses beyond those purely to do with methods of analysis and evaluation, and to try to introduce the topic of ergonomics/human factors from a “doing it” perspective. It is to be hoped that we have succeeded in this aim.

**John R. Wilson**



---

# Acknowledgements to the Third Edition

First of all, I thank all the authors for their great effort in updating chapters or writing new chapters. Particularly for those authors who have been with us since the first edition, I recognise that there is little in it for their careers to publish a revised version of a chapter they have already had in print, and therefore I am extremely grateful for the diligence and care with which everyone has produced revisions. In addition, I thank the anonymous reviewers who made suggestions as to how to improve this third edition and also to all readers and users of the first and second edition who have sent comments and criticisms to the authors and editors.

For this edition, the production of the book has moved from Taylor & Francis Group in the United Kingdom to CRC Press in the United States. It is a measure of the professionalism of all involved that this transition (at the time of writing at least!) appears to have taken place very smoothly. At the publishers, we would like to thank Cindy Carelli, Helena Redshaw and Rachel Tamburri-Saunders for all their help in the production of the book. I also thank Chris Stapleton yet again for her work on the production of the index for the book.

A final thank-you goes to all my colleagues in the Institute for Occupational Ergonomics at the University of Nottingham, and those in other institutions with whom we collaborate. Without working with colleagues who have enthusiasm for the development of new knowledge in ergonomics/human factors and the application of new and existing methods, I would have little enthusiasm to produce books of this nature. I very much welcome the professionalism and friendship of all my colleagues. Finally, I must yet again thank very much my secretary Lynne Mills for all her hard and insightful work on this book, greatly enhancing its production as near to on time as possible and the quality of the content, as well as acting as a buffer between myself and the authors and publishers.

**John R. Wilson**





---

# Preface to the Second Edition

Since the first edition of *Evaluation of Human Work* was published, much has happened to change the way we view ergonomics methods and techniques. This has led to the inclusion of several new chapters in this second edition, and the considerable revision of many others.

Technical, social, political and legal changes have required continual development and improvement of ergonomics methods. For instance, the ever-increasing power and prevalence of computer systems and the diversity of their user interfaces necessitate parallel improvements in methods of analysis, design and evaluation. We can see this, for instance, in human-computer interfaces in general (revised Chapter 12) and in specialised applications such as control rooms (new Chapter 13). Social and political changes, mirrored by changes in the way industrial work and jobs are organised, have increased recognition of the gains possible from greater involvement of people in what they do and from providing employees with a greater degree of control over their own activities. One manifestation of this is participation, and participative approaches (new Chapter 37) have a long and honourable tradition in ergonomics. Legal developments have had a profound influence on ergonomics in recent years, especially the health and safety regulations governing use of display screen equipment, manual handling work, work equipment and workplaces, which have come into force in Europe, Australia and to an extent in North America. Coupled with costs of compensation claims, such regulations have required structured ergonomics assessments at work (new Chapter 30) within an ergonomics management programme (revised Chapter 1 and new Chapter 35). In addition to the above, this edition includes a new chapter on measurement of physiological functions (Chapter 29) and substantial revision of the chapters on task analysis (Chapter 6), verbal protocol analysis (Chapter 7), product assessment and user trials (Chapter 10), knowledge elicitation (Chapter 14), computer aided methods (Chapter 20), mental workload (Chapter 25) and work stress (Chapter 26).

There are increasing moves towards greater professionalisation in ergonomics, for example the Board of Certification in Professional Ergonomics (BCPE) and the Centre for Registration of Ergonomists in Europe (CREE). Such moves require a recognition that the methods we choose will influence what we find from any investigation, and that methods must produce findings that are valid, reliable and generalizable; meet the objectives of the investigation; and are safe and ethical to apply. There can be little excuse for administering questionnaires that make no attempt to use previously validated scales, carry out experiments without careful piloting, or make rigid, ill-informed use of assessment checklists.

Any experimentalist should have a good knowledge of statistics. We decided, after long discussion, not to include statistics in this book – it is heavy enough already! Statistics are necessary not only for experimental design, data compression or testing of results, but for the understanding they bring to the nature of variability and the importance of interactions. Although methods are reported in a ‘stand-alone’ manner, it is rare in ergonomics to find an influence or a cause which has an exclusive relationship with an effect. Hence it is important to retain the ergonomics approach of viewing the whole person within the total environment, an approach which will make it necessary to match a selected group of methods to the requirements perceived. We hope this book will assist ergonomists to do this.



---

# Acknowledgements to the Second Edition

For this second edition we would like to thank all our authors for their considerable efforts in updating their existing chapters or writing new ones; thanks also to reviewers of the new chapters. Many readers of the first edition have sent in comments and criticism to the authors and editors, and these have been accounted for wherever they improved the book.

Again, we must acknowledge the support of Taylor & Francis Group, and especially Richard Steele and Robert Chaundy, and thanks to Chris Stapleton for her professional service on proofreading and the index. The first editor (JW) gratefully acknowledges the Department of Safety Sciences, University of New South Wales, for the space and time to carry out most of the editing and writing of the new chapters.

Finally, we are very grateful to all our colleagues in the University of Nottingham's Department of Manufacturing Engineering and Operations Management, and especially in the Institute for Occupational Ergonomics, for their collaboration, support and a positive environment. Of these, Lynne Mills has contributed the lion's share in terms of typing, editing and organisation – of the book and of us!

**John R. Wilson**  
**Nigel Corlett**



---

# Preface to the First Edition

For a long time there existed few books on ergonomics or human factors methodology; Chapanis' *Research Techniques in Human Engineering*, published in 1959, was probably the earliest, as well as the best known. Lately, there has been a slow increase in what is available. For instance one of the contributors to this volume, David Meister, has produced two books dealing with methods (Meister, 1985, 1986) and the present editors have also been involved in two collections of conference proceedings concerned with new methods and techniques (Laboratory of Industrial and Human Automation, 1987; Wilson et al., 1987).

The books by Meister, excellent in many respects, concentrate upon investigations of large-scale (military) systems design, simulation and evaluation. The two sets of conference proceedings, whilst containing a range of methodological developments and applications, represent what was selected from the papers submitted for presentation, and do not fully represent the field. Also produced recently is the authoritative *Handbook of Human Factors*, edited by Salvendy (1987). This does have much to say about methods and techniques, both as separate chapters or as parts of other chapters; nonetheless its intention is to be a comprehensive, general text, with explanation of theories, principles, data and application, as well as of methods.

Our aim with this book on ergonomics methodology is to produce a text on methods and techniques that is both broad and deep. We intend it to be a companion to the major general textbooks on ergonomics and human factors, particularly and most recently those of Bailey (1982), Grandjean (1988), Kantowitz and Sorkin (1983), Osborne (1987), Salvendy (1987) and Sanders and McCormick (1987). All of these are well known to students, teachers and practitioners of ergonomics, as well as to many of those from other disciplines who take a personal or professional interest in ergonomics. There is, though, little opportunity in such texts to emphasise and make explicit the major part of methodology.

Therefore we have set out to produce a general text on ergonomics methodology. As the book's title implies, we are primarily concerned with people at work and with applied rather than basic research. However, the former concern has not ruled out contributions relevant to people's activities at home, leisure or on the road; nor does the latter concern invalidate descriptions of laboratory-based methods – these can have outcomes that are as practically applicable as are those from field investigations.

The contents of the book are intended to be interesting and useful for a wide range of people, including *students*, to give them a feel for ergonomics investigation and to complement their learning of theory and principles; *industrial and business personal at all levels*, to allow them to understand better what ergonomics can do for them, why and how; and *ergonomics practitioners, researchers and teachers*, to give them a compendium of methods and techniques available. For all these groups, the contributions here will also point to further sources for more detail on specific topics.

Our text on evaluating human work has brought together experts from many branches of ergonomics theory and practice, and has allowed them the space to introduce and give detail on those methods and techniques of value to them. Since ergonomics is both a science and a technology, these methods can of course be concerned with collecting data or with applying their own or others' data. The primary thrust of each contribution may be the general method (e.g. direct observation or protocol analysis), or particular fields of application for several types of methods (e.g. mental workload or the climatic environment). Whilst there will no doubt be omissions – of branches of methodology or of techniques within one area, regretted by some readers – we trust that most will find the book to be a comprehensive, readable and useful source of ergonomics knowledge

and practice. Certainly we believe that for those students or readers from industry who are relatively new to ergonomics, one of the most interesting and valuable ways to learn about it is through its rich and varied methodology.

**John R. Wilson**

**Nigel Corlett**

*University of Nottingham  
Nottingham, United Kingdom*

## REFERENCES

- Bailey, R.W. (1982). *Human Performance Engineering: A Guide for System Designers*. (London, U.K.: Prentice Hall), pp. 656 + xxviii.
- Chapanis, A. (1959). *Research Techniques in Human Engineering*. (Baltimore, MD: John Hopkins Press), pp. 316 + xii.
- Grandjean, E. (1988). *Fitting the Task to the Man: A Textbook of Occupational Ergonomics*, 4th edn. (London, U.K.: Taylor & Francis Group), pp. 363 + ix.
- Kantowicz, B.H. and Sorkin, R.D. (1983). *Human Factors: Understanding People-System Relationships*. (New York: John Wiley & Sons), pp. 699 + xii
- Laboratory of Industrial and Human Automatics (1987). New techniques and ergonomics. In *Proceedings of an International Research Symposium*. (Paris, France: Hermes).
- Meister, D. (1985). *Behavioural Analysis and Measurement Methods*. (Chichester, U.K.: John Wiley & Sons), pp. 509 + ix.
- Meister, D. (1986). *Human Factors Testing and Evaluation*. (Amsterdam, the Netherlands: Elsevier Science), pp. 424 + xi.
- Oborne, D.J. (1987). *Ergonomics at Work*, 2nd edn. (New York: Wiley & Sons), pp. 386 + xvii.
- Salvendy, G. (ed.) (1987). *Handbook of Human Factors* (New York: Wiley & Sons), pp. 1874 + xxiv.
- Sanders, M.S. and McCormick, E.J. (1987). *Human Factors in Engineering and Design*, 6th edn. (New York: McGraw-Hill), pp. 664 + viii.
- Wilson, J.R., Corlett, E.N. and Manenica, I. (1987). *New Methods in Applied Ergonomics*. (London, U.K.: Taylor & Francis Group), pp. 283 + x.

---

# Acknowledgements to the First Edition

Our first debt with this book is to our contributing authors, all of whom have responded to our various requests with great patience, and have produced chapters of high quality within, in some cases, a very limited time. Amongst these authors we must mention those who were with us in the initial discussions about the book at the Second International Occupational Ergonomics Symposium at Zadar, Yugoslavia; they were Lisanne Bainbridge, Colin Drury, Ted Megaw, Ken Parsons, Pat Shipley and Rob Stammers. Colin Drury in particular has contributed much in terms of individual chapters, and the overall content and style of the book.

We thank our colleagues at Nottingham University for contributing to a working environment in which we feel able to embark on and complete this and other publishing ventures. One of us (JW) also thank the Department of Industrial Engineering and Operations Research, University of California, Berkeley, for allowing him time and facilities to work on this book during periods there as a visitor in 1987 and 1988.

Our editors at Taylor & Francis Group – David Grist, Sarah Waddell and, for most of the time, Robin Mellors – have been exceedingly supportive, even in the face of a project which seemed to grow exponentially! The style of the book has been enhanced tremendously by the artwork of Tony Aston and the cartoons by Moira Tracy. Despite both editors being away from Nottingham for substantial periods of time, the publication of their book has rolled on relatively smoothly; our colleagues would say this was because we left this and much else in the hands of our excellent secretaries, Lynne Mills and Ilse Browne, to whom we are immensely grateful.





---

# Editors

**John R. Wilson (Deceased)** was a professor of human factors at the University of Nottingham, Nottingham, United Kingdom, and also worked for many years part time as a human factors specialist within Network Rail, the U.K. rail infrastructure provider. He was both a chartered psychologist and a chartered engineer, a fellow of the Institute of Ergonomics and Human Factors (president for 2010–2011) and a member of the HFES and Registered European Ergonomist. He was editor-in-chief of *Applied Ergonomics*. In 1995, he was awarded the Sir Frederic Bartlett Medal of the Ergonomics Society for services to international ergonomics teaching and research and was awarded the Distinguished Overseas Colleague Award of the U.S. Human Factors and Ergonomics Society in 2008. He served for many years on the Council and the Executive of the International Ergonomics Association.

**Sarah Sharples** is a professor of human factors at the University of Nottingham, Nottingham, United Kingdom. She leads a number of projects investigating novel technologies in workplace contexts, including transport, manufacturing and healthcare, collaborating with a wide range of industry and academic partners. She is a fellow of the Institute of Ergonomics and Human Factors and will be president of the Institute in 2014–2015. She led the Human Factors Research Group at the University of Nottingham to the award of the President's Medal from the Institute of Ergonomics and Human Factors in 2013.



---

# Contributors

**Chris Baber**

School of Electronic, Electrical and Systems  
Engineering  
University of Birmingham  
Birmingham, United Kingdom

**Nora Balfe**

Department of Psychology  
Centre for Innovative Human Systems  
Trinity College Dublin  
Dublin, Ireland

**Ann Bisantz**

Department of Industrial and Systems  
Engineering  
University at Buffalo  
The State University of New York  
New York, New York

**Tiffany Bisbey**

Institute of Simulation and Training  
University of Central Florida  
Orlando, Florida

**Peter Buckle**

The Helen Hamlyn Centre for Design  
Royal College of Art  
London, United Kingdom

**Gary Burnett**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Rebecca Charles**

School of Aerospace, Transport  
and Manufacturing  
Cranfield University  
Bedfordshire, United Kingdom

**Young-Jin Cho**

ErgoMechanics Laboratory  
Hanyang University  
Ansan, South Korea

**Sue Cobb**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Tom Cox**

Centre for Sustainable Working Life  
School of Business, Economics and Informatics  
Birkbeck University of London  
London, United Kingdom

**Patrick G. Dempsey**

Office of Mine Safety and Health Research  
National Institute for Occupational Safety  
and Health  
Pittsburgh, Pennsylvania

**Colin Drury**

Department of Industrial and Systems  
Engineering  
University at Buffalo  
The State University of New York  
Buffalo, New York

**Clive D'Souza**

Department of Industrial and Operations  
Engineering  
University of Michigan  
Ann Arbor, Michigan

**Ken Eason**

Loughborough Design School  
Loughborough University  
Loughborough, United Kingdom

**Jörgen A.E. Eklund**

Unit of Ergonomics  
School of Technology and Health  
KTH Royal Institute of Technology  
Stockholm, Sweden

**Edward Elton**

School of Computing, Engineering  
and Mathematics  
University of Brighton  
Brighton, United Kingdom

**David Feathers**

College of Human Ecology  
Cornell University  
Ithaca, New York

**Mike Foley**

Washington State Department of Labor  
& Industries  
Tumwater, Washington

**Jane Fulton Suri**

IDEO  
The Embarcadero  
San Francisco, California

**David Golightly**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Mike Gray**

Health and Safety Executive  
Sheffield, United Kingdom

**Amanda Griffiths**

Division of Psychiatry and Applied  
Psychology  
School of Medicine  
University of Nottingham  
Nottingham, United Kingdom

**Diane Gyi**

Loughborough Design School  
Loughborough University  
Loughborough, United Kingdom

**Christine M. Haslegrave**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Sue Hignett**

Loughborough Design School  
Loughborough University  
Loughborough, United Kingdom

**Robert Houghton**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Peter A. Howarth**

Loughborough Design School  
Loughborough University  
Loughborough, United Kingdom

**Harrison Kelly**

Department of Industrial and Systems  
Engineering  
University at Buffalo  
The State University of New York  
Buffalo, New York

**Jung-Yong Kim**

ErgoMechanics Laboratory  
Hanyang University  
Ansan, South Korea

**David Kirk**

School of Computing Science  
Newcastle University  
Newcastle upon Tyne, United Kingdom

**Barry Kirwan**

Eurocontrol Experimental Centre  
Bretigny sur Orge, France

**Kenneth R. Laughery**

Department of Psychology  
Rice University  
Houston, Texas

**Glyn Lawson**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Alyssa Marshall**

Institute of Simulation and Training  
University of Central Florida  
Orlando, Florida

**Christopher B. Mayhorn**

Department of Psychology  
North Carolina State University  
Raleigh, North Carolina

**Ian McClelland**

Philips Digital Systems Laboratory  
Eindhoven, Netherlands

**Hilary McDermott**

School of Sport, Exercise and Health Sciences  
Loughborough University  
Loughborough, United Kingdom

**Ted Megaw**

School of Electronic, Electrical and Computing  
Engineering  
University of Birmingham  
Birmingham, United Kingdom

**Ranjana Mehta**

Department of Environmental  
and Occupational Health  
Texas A&M University  
College Station, Texas

**Seung-Nam Min**

ErgoMechanics Laboratory  
Hanyang University  
Ansan, South Korea

**Moirra Munro**

Glasgow, United Kingdom

**Colette Nicolle**

Loughborough Design School  
Loughborough University  
Loughborough, United Kingdom

**Victor Paquet**

Department of Industrial and Systems  
Engineering  
University at Buffalo  
The State University of New York  
Buffalo, New York

**Raja Parasuraman**

Department of Psychology  
George Mason University  
Fairfax, Virginia

**Ken Parsons**

Loughborough Design School  
Loughborough University  
Loughborough, United Kingdom

**Jonisha P. Pollard**

Office of Mine Safety and Health Research  
National Institute for Occupational Safety  
and Health  
Pittsburgh, Pennsylvania

**Jane A. Rajan (Deceased)**

Rice University  
Hertfordshire, United Kingdom

**Emilie M. Roth**

Roth Cognitive Engineering  
Menlo Park, California

**Brendan Ryan**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Eduardo Salas**

Institute of Simulation and Training  
University of Central Florida  
Orlando, Florida

**Nigel R. Shadbolt**

School of Electronics and Computer Science  
University of Southampton  
Southampton, United Kingdom

**Sally Shalloe**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Sarah Sharples**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Andrew Shepherd**

Loughborough, United Kingdom

**Carys E. Siemieniuch**

School of Electrical, Electronic and Systems  
Engineering  
Loughborough University  
Loughborough, United Kingdom

**Murray A. Sinclair**

School of Electrical, Electronic and Systems  
Engineering  
Loughborough University  
Loughborough, United Kingdom

**Gisela Sjøgaard**

Institute of Sports Science and Clinical  
Biomechanics  
University of Southern Denmark  
Odense, Denmark

**Paul R. Smart**

School of Electronics and Computer Science  
University of Southampton  
Southampton, United Kingdom

**Karen Sjøgaard**

Institute of Sports Science and Clinical  
Biomechanics  
University of Southern Denmark  
Odense, Denmark

**Tom Stewart**

System Concepts  
London, United Kingdom

**Emile Tompa**

Institute for Work and Health  
Toronto, Ontario, Canada

**Dwayne Van Eerd**

Institute for Work and Health  
Toronto, Ontario, Canada

**Patrick Waterson**

Loughborough Design School  
Loughborough University  
Loughborough, United Kingdom

**Jennifer Watts-Englert**

PARC, a Xerox Company  
New York, New York

**John R. Wilson (Deceased)**

Human Factors Research Group  
Faculty of Engineering  
University of Nottingham  
Nottingham, United Kingdom

**Michael S. Wogalter**

Department of Psychology  
North Carolina State University  
Raleigh, North Carolina

**John Wood**

CCD Design and Ergonomics Ltd.  
London, United Kingdom

**Paul H.P. Yeow**

School of Business  
Monash University Malaysia  
Selangor, Malaysia

---

# 1 Methods in the Understanding of Human Factors

*John R. Wilson and Sarah Sharples*

## CONTENTS

|                                                                                  |    |
|----------------------------------------------------------------------------------|----|
| Introduction.....                                                                | 1  |
| Some Historical Background .....                                                 | 5  |
| Definitions of Ergonomics/Human Factors.....                                     | 6  |
| Aims and Framework for Ergonomics.....                                           | 8  |
| Argument for and against E/HF .....                                              | 10 |
| Context for Application of E/HF Methods.....                                     | 13 |
| Role of E/HF Methods.....                                                        | 15 |
| Methods for the Collection of Data about People and Their Needs .....            | 15 |
| Methods Used in Systems Analysis and Design.....                                 | 15 |
| Methods to Evaluate Human–Machine System Performance.....                        | 16 |
| Methods to Assess Demands and Effects on People.....                             | 16 |
| Methods to Develop E/HF Management Programmes .....                              | 17 |
| Classification of Methods .....                                                  | 17 |
| Debates and Discussions around Methodological Approaches .....                   | 18 |
| Qualitative/Quantitative Debate for Methodology in Ergonomics/Human Factors..... | 22 |
| ‘Objective’ and ‘Subjective’ Methods.....                                        | 22 |
| Field and Laboratory Study.....                                                  | 24 |
| Quality of Methods and Measures and Their Choice .....                           | 25 |
| Validity .....                                                                   | 26 |
| Reliability .....                                                                | 26 |
| Generalisability .....                                                           | 27 |
| Interference .....                                                               | 27 |
| Sensitivity.....                                                                 | 27 |
| Feasibility of Use .....                                                         | 27 |
| Acceptability and Ethics .....                                                   | 27 |
| Resources .....                                                                  | 28 |
| Conclusions.....                                                                 | 28 |
| References.....                                                                  | 29 |

## INTRODUCTION

It is common to introduce the idea of ergonomics, and of the importance of the human factor, by referring to one of a number of well-publicised disasters in which major loss of life and great commercial costs occurred – from Flixborough to Bhopal, from Three Mile Island to Chernobyl, from Piper Alpha to Ladbroke Grove. Many examples can be found, for instance, in Beaty (1995), Bignell et al. (1977), Casey (1993), Cushing (1994), Perrow (1984), Petroski (1992), Reason (1990) and Proctor and Zandt (2008). Recent examples include *Deepwater Horizon*, the aftermath of

Hurricane Katrina, the *Costa Concordia* capsize and the Santiago derailment and demonstrate the continued need to understand the role of design and the human factor in the management of safety and performance.

An incident that led to significant embarrassment for the Submarine Service within the British Royal Navy, but fortunately not loss of life, was the grounding of the nuclear-powered submarine, HMS *Astute*, off the west coast of Scotland in 2010. The official report into the incident revealed a number of contributing human factors, including: limited experience and lack of familiarity of the officer of the watch with the geographic location (the commanding officer was unaware of this deficit in geographical knowledge); lack of a dedicated plan or specific briefing for the upcoming transfer; absence of key members of the team (and a lack of awareness of the implications of this absence on team expertise); late-running of the submarine for a rendezvous (and the impact of this on choice of actions); missing equipment on the bridge within the submarine; faulty communications equipment and the primary radar not being used, because it was emitting noise (despite the fact that this 'did not affect the serviceability of the radar') (UK Government, 2012). The senior commanding officer (who was away from the bridge at the start of the incident) lost his post after the incident, but the fact that of the ten recommendations in the official report, four relate to personnel failure, two relate to process and four relate to equipment quality (including one specifying a need for better monitoring equipment to capture data in the case of such incidents) indicates that we are perhaps starting to move beyond the assumption that any incident that involves human factors should be wholly attributed to 'human error'.

Across a whole range of ICT (information and communications technology) implementation projects, we can find examples where: end users are not clearly identified and their needs not accounted for; where systems become excessively complex or have far more functionality than really needed; user training and support are not sufficient; interfaces hinder rather than help users; and the development process itself lacks clear objectives, is inflexible and is highly technology driven (e.g. Eason, 1997; UK National Audit Office, 1999; UK Public Accounts Committee, 2000). A very well known example, which broke most of the 'rules' of successful ergonomics, was the failed implementation of a new control system for the London Ambulance Service, which resulted in the loss of life and great financial costs (e.g. Wastell and Cooper, 1996; Beynon-Davies, 1999). The failure of the implementation of the FiReControl project in the United Kingdom (National Audit Office, 2011) wasted £469 million government money, attributed to 'underestimation of complexity of risk' of the project and the attempt to 'centrally impose a national control system on unwilling locally accountable bodies'.

We have in the past perhaps been guilty as a community of focussing on such examples of negative consequences to make the 'case' for E/HF. But the discipline has now existed for over 60 years, and E/HF has been implemented with much success in many domains. For example, the US Department of Labor maintains a list of 'success stories'\* demonstrating the positive impact of E/HF interventions on employee health and performance. Identifying successful E/HF requires us to 'prove the null hypothesis' – demonstrating that an effect is absent, rather than present. We hear much less about air incidents, such as the fire after landing of Air France Flight 358 in Toronto in 2005, where all passengers successfully evacuated within 90s, than we do about an incident with a less positive outcome, yet it is highly likely that the successful evacuation was at least partly enabled by E/HF work by researchers such as Helen Muir, who conducted an extensive programme of research on aircraft evacuation after the Manchester aircraft Flight 28M fire in 1985 (Muir, 1996). The approach of focussing on when things go right, rather than when things go wrong, is embodied in the Safety II approach (e.g. Hollnagel, 2014).

Yet many examples of inappropriate design of equipment, workplaces, systems, jobs and organisations can still be found in large and small companies, in offices and factories and in physical and mental work. The common denominator in all cases is that the abilities, needs and limitations of the people working in the system or with the equipment have not been understood and accounted for in

---

\* [https://www.osha.gov/SLTC/ergonomics/success\\_stories.html](https://www.osha.gov/SLTC/ergonomics/success_stories.html).



the context of the demands of the job. On the other hand, successful products or work systems will usually show evidence that the needs of their users have been accounted for during design, implementation and operation.

Taking such account of people – or of the human factor – is the province of *ergonomics* and of the synonymous *human factors*. (The interchangeability of the terms is indicated in this chapter through use of the shorthand E/HF). This book is about the *methods* that ergonomists and human factors engineers and scientists use – in analysing, designing and evaluating equipment, tasks, jobs and organisations. Ergonomics has little value unless it is applied, and so its practical methodology is of great importance (see Dul et al., 2012 for further discussion of the ‘value proposition’ of E/HF).

The consequences of not applying E/HF, or of wrongly applying E/HF through inappropriate methodology, can lead to increased risk of ill-health and injury, dissatisfaction and discomfort for the workforce. For a company, the consequences at the least can be a loss of competitiveness, in terms of productivity, quality, flexibility and timeliness, as well as a loss of trust in the company, by either the public or its government. However, this book is even more concerned with the positive side of applying E/HF methodology, with the improvements in well-being that can result for employers, workforces, producers, users, engineers, designers and, indeed, people and society in general.

This application of methodology will be in both of what Kragt (1992) – in the context of manufacturing industry – distinguished as *product ergonomics* and *production ergonomics*, denoting the importance of considering design of products, objects or artefacts, as well as developing tools and methods to support the design of processes to design, produce and deliver these items. However, E/HF applies more widely than in products or production, increasingly focussing on service provision and systems. Goods and services are planned and designed by people, produced by people with processes built and managed by people, and are sold, bought, used, maintained and scrapped/recycled by people. Table 1.1 illustrates how E/HF might have a role at different stages in the traditional product life cycle. Even where automated processes are employed, these interface with people and organisational systems at some point, and once a product or system is purchased or implemented, the focus of effort may well be on maintaining and supporting a ‘service’ relationship with the consumer. Therefore, understanding of people, and using that understanding wisely, are central to the planning and running of successful organisations, communities and societies.

However, E/HF has an interesting challenge – its impact is often long term, rather than short term, therefore can be hard to assess on an instantaneous basis. The change from managerial capitalism to shareholder capitalism has forced companies to ignore long-term consequences to concentrate on this quarter’s financial results, despite many more worthy aspirations in company documents. Even the continuous election cycle mitigates against long-term thinking.

This challenge also affects how we select, apply and interpret methods. For example, Mack and Sharples (2009) demonstrated that there were some differences between the extent to which people *thought* that E/HF (in this case, the perceived usability of a mobile phone) was important, compared to whether usability actually influenced their product choice. So increasingly perhaps, the role of the ergonomist is not only to enhance consideration of E/HF at all stages of the product life cycle, but also to clearly demonstrate to the user, purchaser or decision maker that this has been done, and what the improvement in product quality or system effectiveness was as a result.

Whatever the original roots of E/HF, as our home and work boundaries become increasingly blurred, it is now certainly not limited to the workplace, nor to ‘the operator’. E/HF is relevant to all purposive human behaviour in ‘designed’ human–machine systems. In addition, increasingly the ability of an individual to customise, adapt and design their own work methods and workplace setup means we are unable to ‘predict’ the way in which people complete their work tasks. Therefore, E/HF has a role to move well ‘beyond the workplace’ and ensure not only that products are designed for use in a range of contexts, but that our methods consider the variety of ways in which work tasks might be completed. This means that we have to very carefully consider aspects of methods such as the generalisability and transferability of results.

**TABLE 1.1**  
**Role of Ergonomics and Human Factors throughout the Product Life Cycle**

| Stage in Product Life Cycle | Example Role of E/HF                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Concept idea generation     | Support creativity in idea generation<br>Move beyond ‘design blindness’ caused by familiarity with previous products/design solutions<br>Represent different stakeholder views<br>Present guidelines (e.g. physical, cognitive capabilities; environmental considerations) in usable form to enable early consideration                                                                                                                                                                                                                                                                                                                                                                    |
| Design                      | Enable collaboration and participation in design<br>Provide tools to enable designers to appreciate user capabilities and preferences<br>Support effective elicitation and presentation of user requirements<br>Ensure monitoring and updating of user requirements through multiple iterations of design                                                                                                                                                                                                                                                                                                                                                                                  |
| Prototyping                 | Ensure selection of appropriate type of visualisation for stage in design process (early/late iterations) (e.g. using virtual reality, paper prototypes, physical mock-ups)<br>Provide benchmarking against other parameters or data (e.g. a tool to enable workload of proposed system to be predicted, or a digital human modelling system that is used to measure range of population that will be accommodated by a system)<br>Conduct user evaluations with prototypes to anticipate impact of changes on user and workplace performance, both by considering new product in isolation, and interaction between new product and other systems/artefacts in anticipated context of use |
| Manufacture                 | Design safe, comfortable and efficient manufacturing workplaces<br>Identify points in production where errors could occur, and introduce elements to system to improve resilience<br>Consider impact of automation in workplace, and understand how new automated and intelligent workplaces can combine with human operator’s capabilities                                                                                                                                                                                                                                                                                                                                                |
| Distribution and sale       | Ensure prospective user is able to anticipate experience of product use as accurately as possible (e.g. allow prospective purchaser of a new car to visualise product quality from virtual representation)<br>Clearly communicate product capability via product design                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Use                         | Evaluate systems and tools <i>in situ</i><br>Collect opinions on use from users<br>Measure impact of changes on operator/workplace performance                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Maintenance                 | Design products to be ‘maintainable’ to minimise impact on overall work system<br>Clearly communicate to the user which parts of a product or system are maintainable by the user themselves                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Disposal/recycling          | Label items and materials clearly to ensure appropriate recycling/disposal behaviours<br>Make products easy to dismantle to make disposal/recycling as easy as possible                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

*Note:* Example activities are shown in the table, this is not intended to be an exhaustive list.

Thus, E/HF is highly relevant to, and for, people at home, on the move, in sport, at leisure as well as at work, and for school children, post-retirement populations and those with physical and cognitive impairments. Even at work, it is probable that a minority of people are working in jobs traditionally categorised as ‘operators’ (usually assembly line staff, machine operators or process control operators). People at work have a vast variety of roles, interact with a wide range of products and systems, have multiple goals and means and require integration of social, cognitive and physical skills, particularly in the service industries, where the customer also has a key role in the task.

## SOME HISTORICAL BACKGROUND

Clarity of focus for E/HF has been bedevilled by the use of the different terms for the discipline itself. *Ergonomics* and *Human Factors* are now synonymous and accepted terms worldwide for the theory and practice of learning about human characteristics and capabilities, and then using that understanding to improve people's interaction with the things they use and with the environments in which they do so. The interchangeable use of both terms is demonstrated by the change in the name of the UK 'Ergonomics Society' to the 'Institute of Ergonomics and Human Factors' in 2009.

Distinctions are increasingly blurred. One difference is that people from many different disciplines will say they work in 'human factors' but generally only those with a degree in 'ergonomics' or, confusingly, 'human factors', will say they work in 'ergonomics'. There is a tendency, even amongst ergonomists, to talk about 'human (or people) factors' as being what they study, simply because this is easier grammatically in written and spoken language! The converse is that people working in the profession of human factors find it difficult to know what to call themselves, with 'human factors engineers' 'human factors scientists' and 'human factors professionals' all having been tried. Other efforts to get around the problem of name – for instance, using the title of engineering psychology – tend to lose the holistic and total systems flavour of E/HF. Interestingly, the current term used in many US bodies, such as the National Research Council, is 'Human Systems Integration' – a term that emphasises the systems approach, but perhaps distances itself from the underpinning *discipline* of E/HF.

The ease of use of term 'human factors' does have some unintended consequences though. It is increasingly being adopted by specific domain areas, thus we come across terms such as 'clinical human factors' or 'nuclear human factors'. This approach can result in a specific focus on a particular aspect of E/HF that is particularly relevant to the domain (anecdotally, in the safety critical context of healthcare, the focus of activity is often on the role of human communication, teamwork and collaboration for example). This may simply be a result of the strong influence of specific and high profile cases that have rightly emphasised the importance of human factors within these areas, but it is important that the breadth and systems approach of E/HF is not lost. The increasing profile of human factors within specific domains (particularly the healthcare domain, where there are many activities in Europe, US and UK to work with clinical professionals to increase consideration of E/HF in many types of clinical settings) is an excellent thing, and brings the active engagement not only of expert ergonomists but also clinical professionals, whose buy in is essential to ensure deep embedding of E/HF approaches within the clinical setting. This perhaps demonstrates the importance of us embracing the meaningfulness of the term 'human factors' to colleagues outside the core discipline, and making sure that the tools and methods we develop are usable by those who may not be 'trained ergonomists' but who do have vital familiarity with and tacit knowledge of their own workplace setting.

We may regret that the term 'ergonomics' lacks innate meaning or impact for clients or the public who will not be aware of its classical roots, but genies can rarely be put back into bottles; it would now be a futile and damaging exercise to move away from use of the terms ergonomist, ergonomics and human factors. In this book, we emphasise that it is what we do that matters, not what we call ourselves, and contributors will use the terms ergonomics and human factors interchangeably. We must work in teams as systems engineers, human factors integration specialists, capability and performance managers, to embed E/HF tools and practice within different industrial contexts.\*

Although formal consideration of the interactions between people and their working environments can be found in writings of a hundred years ago, for instance, from Poland and Germany, the modern history of E/HF emerges from the 1939 to 1945 war. As a formal branch of learning,

---

\* I have often noted that I practise 'ergonomics by stealth' (Colin Drury uses the term 'guerrilla ergonomics') – sometimes by simply being part of a team, giving anecdotal examples of the impact of ergonomics over a coffee or lunch break, or just giving a different perspective on a design suggestion, can be an important part of encouraging colleagues from other disciplines to take E/HF into account (SS).

with its own learned societies and scientific journals, E/HF has a formal history of about 60 years, for instance in Germany, the Netherlands, the United Kingdom and the United States. In the United Kingdom, the ideas and expertise from different disciplines interested in the effectiveness of human performance (anatomy, physiology, industrial medicine, industrial hygiene, design engineering, architecture and illumination engineering), and an emphasis on theory and methodology, led to the formation of the discipline of ergonomics with two strong sub-groupings: those of anatomy/physiology and experimental psychology. In parallel, the human factors profession was growing up in the United States, with strong inputs from the disciplines of psychology and engineering. In Germany, the Netherlands and across Scandinavia, a basis for ergonomics was growing out of work in medicine and functional anatomy, whilst in Eastern Europe, growth was largely from the industrial engineering profession. For much more background detail, interested readers are referred to Edholm and Murrell (1973), Singleton (1982), Stockbridge (1989) and Meister (1995b) and histories by Waterson and Sell (2006) and Waterson (2011) for the Institute of Ergonomics and Human Factors and by Kuorinka (2000) for the International Ergonomics Association.

E/HF has drawn from anatomy, physiology and psychology, and has close connections with the applied disciplines of medicine and engineering. Chapanis (1996) defines E/HF as a multidisciplinary field, with psychology (primarily experimental psychology), anthropometry, applied physiology, environmental medicine, engineering, statistics, operations research and industrial design all contributing. Wickens et al. (2013) describe how the field of human factors grew from initial considerations of performance in military contexts, but has broadened greatly in the last few decades with its various sub-domains and application contexts (they note consumer products, business, highway safety, telecommunications and healthcare). Overlapping and related disciplines and names, often influenced by academic cultures as much as linguistic meanings of terms, include experimental psychology, engineering psychology, social psychology, industrial engineering, bioengineering, cognitive science, artificial intelligence, systems engineering, human-computer interaction, interaction design, industrial design, management and statistics.

## DEFINITIONS OF ERGONOMICS/HUMAN FACTORS

It is a feature of the modern world that disciplines of current relevance and value are generally multi-, inter- and trans-disciplinary and therefore less amenable to simple definition. The fact that E/HF was built upon existing fundamental disciplines should not in itself be a problem: engineering is built upon mathematics and physics, psychology is built upon biology and economic science upon a number of bases.

E/HF seeks to define itself at regular intervals – for instance, amongst many others, see Chapanis (1976, 1979), Welford (1976), de Montmollin (1992), Moray (1994) and Meister (1998). A large number of different, if overlapping, definitions of ergonomics and of human factors now exist; Wogalter et al. (1998) considered these. Most definitions stress the view of E/HF as jointly a science – providing fundamental information – and a technology – applying that information to problems of design in their widest sense (e.g. Shackel, 1996). Within this view, the E/HF sphere contains all elements of the total human–environment system, comprising people’s interactions with hardware, software and ‘firmware’ (including space), and with other people (‘liveware’) both individually and as social groups.

Any acceptable definition of E/HF must emphasise the need for, and the complementarity between, fundamental understanding of people and their interactions and practice in improving those interactions. Meister (1995a, p. 9) differentiates between the theoretical knowledge within ergonomics, which explains people’s interaction with other things, and the instrumental knowledge, which can be utilised in design. This relationship between theory and practice, between research and application, is under examination continually (e.g. Singleton, 1994; Green and Jordan, 1999). A US National Research Council report (Rouse et al., 1997) makes strong arguments for the value

of human factors initiatives in many industries, stressing its multidisciplinary, systemic, sociotechnical and user-centric orientation, but also the difficulties in the current climate of balancing needs for basic research with the high demands for applied activities. As Wilson (2000) notes ‘we can see ergonomics as comprising elements of craft, science and engineering; it has aims to implement and evaluate (craft), to explain and predict (science), and to design for improved performance (engineering)’.

The International Standards Organization, in its various committees on ergonomics standards, has been using as a working definition that:

Ergonomics produces and integrates knowledge from the human sciences to match jobs, systems, products and environments to the physical and mental abilities and limitations of people. In doing so it seeks to safeguard safety, health and well-being whilst optimising efficiency and performance.

Similarly, the International Ergonomics Association (as of 2014) has defined that:

Ergonomics (or human factors) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theoretical principles, data and methods to design in order to optimise human well-being and overall system performance.

It is worth highlighting the final phrase *optimise human well-being and overall system performance* – it could be argued that this is really the ‘charter’ for this book, and that our goal is to measure all aspects of human well-being and system performance. The definition continues to explain:

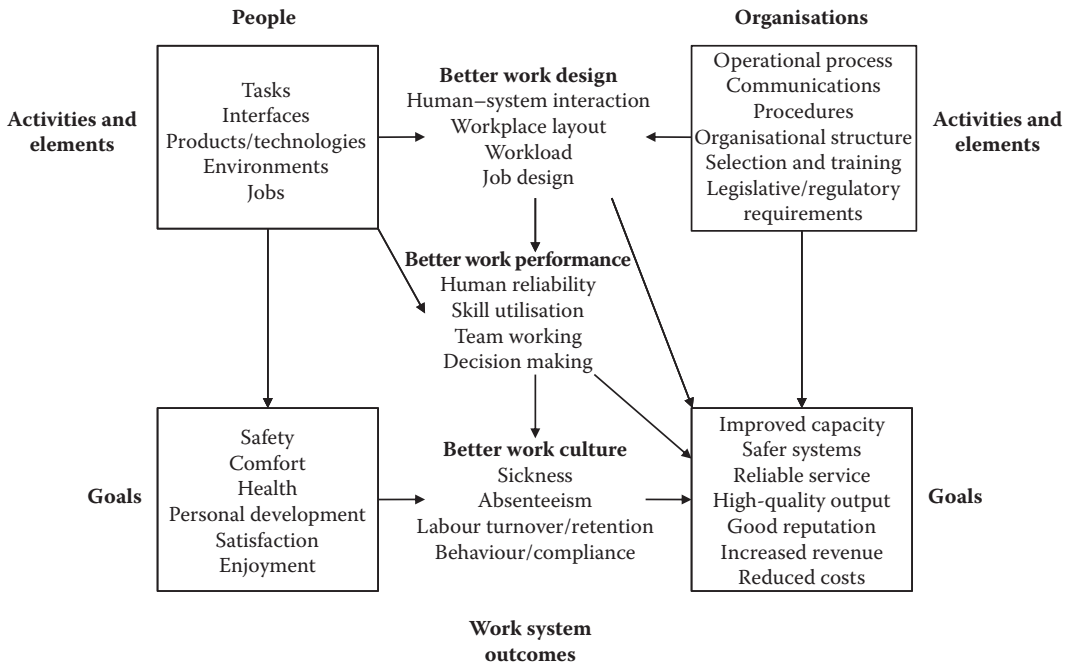
Practitioners of ergonomics and ergonomists contribute to the design and evaluation of tasks, jobs, products, environments and systems in order to make them compatible with the needs, abilities and limitations of people. Ergonomics helps harmonise things that interact with people in terms of people’s needs, abilities and limitations.

When pushed at a party for rather more pithy definitions of ergonomics, many of us will respond that it is design for people, or designing fit for a range of human users, or fitting systems and products to people and not *vice versa* (Figure 1.1).

A prevalent current tendency to define and partition the work of ergonomists into specialisms, whilst understandable, can cause difficulties. Typically this is done into cognitive ergonomics, physical (or musculoskeletal) ergonomics and social (or organisational) ergonomics; other specialisations are also sometimes defined such as rehabilitation ergonomics, green ergonomics and forensic ergonomics. There are some good reasons for such partition. It can help explain a potential contribution to clients and funding bodies: for instance, an ergonomist specialising in complex control systems design may define what they practise as cognitive ergonomics; an ergonomist specialising in health and safety at work and re-design of workstations may define this as physical ergonomics. Nonetheless, there are dangers in such a parochial and molecular viewpoint. It is the very systems perspective and holistic nature of E/HF that provides its strength. The breadth of concern to cover all aspects of people’s interaction with their environments and the interconnections between these interactions is what allows E/HF to define itself as a unique discipline, and is at the heart of the certification or chartership of many professional bodies associated with the discipline, such as CREE (Centre for Registration of European Ergonomists), the BCPE (Board of Certification of Professional Ergonomists) and the IEHF (Institute of Ergonomics and Human Factors).

If we do see the value in descriptions of specialisms, these might include:

- *Physical ergonomics*: fit, clearance, reach, access, tolerance, workload, manual handling, health and safety, workplace layout, displays and controls, product and equipment design, environment, tools



**FIGURE 1.1** Aims of ergonomics/human factors. E/HF can be seen in the context of its objectives, which are the well-being of people and of organisations. These might be seen as twin aims, which are neither independent nor mutually exclusive and which have direct and systemic connections.

- *Cognitive ergonomics*: information processing, sensing, perception, decision making, problem solving, reaction, mental workload, fatigue, stress, interface design, reliability, communication, fault diagnosis
- *Organisational (social) ergonomics*: attitudes, motivation, satisfaction, job and team design, hours and patterns of work, pacing, implementation of change
- *Systems ergonomics*: taking a holistic approach to design and evaluation that integrates the physical, cognitive and organisational/social

## AIMS AND FRAMEWORK FOR ERGONOMICS

In most definitions of E/HF, we will find a list of objectives or criteria that drive its application, for instance jobs, systems or products that are comfortable, safe, effective and satisfying. Aims are often divided into those of gains for the individual (employee or user), and those for the organisation (employer or producer). *These aims, however, are neither mutually exclusive nor independent.* It is not a case of having to implement either a more comfortable workstation or a more productive one for example, nor are the ways of achieving the former necessarily very different from those for the latter, and many (e.g. Becker and Gerhart, 1996) have demonstrated the positive impact of employee well-being on organisational performance. In fact, it is difficult to think of cases where work or equipment designed to meet the needs of an employee or user would detract from performance effectiveness; in general, user fit will enhance performance. To contribute such ‘win-win’ outcomes, and to sell them before and to prove them afterwards to managers and engineers, is one of the main tasks facing practising ergonomists. Generally, if we concentrate upon objectives of ergonomics, then we require methods that go some way to helping us ‘prove’ that we have met certain aims or have achieved a certain level of improvement (see Kim et al., 2001 for an example that demonstrates this approach in an assembly line context). Cost–benefit analyses (see Chapter 35), an



appreciation of the relationship between ergonomics and quality (see Chapter 36) and the setting of usability metrics (see Chapter 13), are of importance here.

E/HF remains of importance in the face of the changing nature of work. As mentioned earlier, the home/work boundary is becoming blurred. Increasingly, we use our home technologies (e.g. personal mobile phones) to support our work, and our work technologies to support our home activities (e.g. accessing a website at work (during a lunch hour of course!) to organise our online supermarket shopping delivery). It would be foolish to ignore these changes. Our expertise in using work systems can be enhanced by our home practice – there is an argument for designing the interface for a system used in a workplace to control a tele-operated robot to follow interaction paradigms established by a personal games system, for example. In addition, as has been discussed extensively by Erik Hollnagel and David Woods (2005), it is no longer appropriate to consider the person and the technology/system as two separate elements, with inputs and outputs between them; instead, it can be more fruitful to consider how people work *in partnership* with technology as a ‘joint cognitive system’ (this approach also has synergies with concepts of distributed cognition [Hutchins, 1995] and concepts such as computational offloading [Scaife and Rogers, 1996]).

Figure 1.1 illustrates the twin aims of E/HF in the context of work systems. We see that there is a direct connection between design and development criteria for people and organisations, and an indirect or systemic one also. Demonstrating the value of considering both people *and* organisations as contributors to better work systems strengthens the argument for the consideration of E/HF.

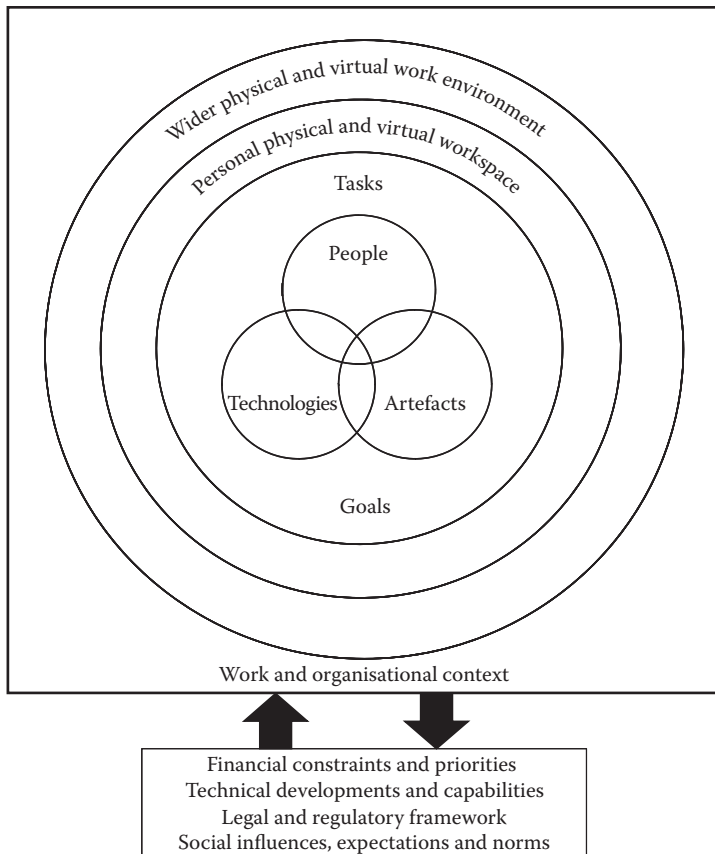
A traditional view is that, like the epidemiological model of ‘host–agent–environment’ in disease control or accident prevention, E/HF is concerned with interactions between people and the things they use and the environments in which they use them. Our concern is with the contextually based (or situated) interfaces between people and the processes with which they interact, whether a toothbrush, training manual, motor car, power plant control room or school, for instance. Within this, people’s interaction with other people is at the core of an E/HF framework, increasingly so as systems of interest move away from ‘one person – one interface’ to distributed networks (Wilson, 2000).

Traditionally, we have represented the context of E/HF using the ‘onion model’ (Grey et al., 1987), now a classic representation of the ‘person at the centre’ and used in many introductory classes to set the scene of E/HF. Figure 1.2 shows the ‘onion revisited’ – a revised look at the interactions between factors relevant to the application of E/HF in work design.

The centre shows the three overlapping elements representing ‘individual interaction’ – a person, their technologies and artefacts that they are using. If we consider an example of the Air Traffic Controller (ATCO), they would work with, at the very least, communications technologies and a representation of aircraft location (usually via a radar system of some sort), will use paper or virtual flight strips to record communications and instructions and will be in a room/at a desk where artefacts include clocks, diagrams of airspace or schedule information. The ATCO will complete discrete tasks (e.g. approve a flight level change request) and have goals (e.g. maintain safe separation of aircraft, ensure planes depart according to schedule) which influence their completion of those tasks and their interactions with technologies and artefacts.

Task completion takes place in a personal physical and virtual workplace – we need to move ‘beyond the desk’ when we consider a person’s workplace. Increasingly we store our personal documents and information in the ‘cloud’ – a distributed network of servers that enables us to access stored data resources remotely, regardless of our physical location. Although this may be in a physically different place to an individual, it is still their ‘personal’ space in this virtual system. We should not of course overlook the continued importance of local workspace design (e.g. see Chapter 14 on control facilities design for further consideration of this matter).

Similarly, the wider work environment takes on a physical and virtual form. Environmental considerations remain a significant contributor to workplace performance and satisfaction, and are increasingly automated and ‘intelligent’ (and we have all sat in freezing cold ‘intelligent’ environments, wondering how we can break into the air conditioning system), presenting challenges



**FIGURE 1.2** The 'onion model' revisited – Interactions of factors relevant to the application of E/HF in work design. At the centre is the consideration of people, technologies and artefacts. These are placed in the context of their tasks and goals, physical and virtual workspaces and organisational context, and are influenced by financial, technical, legal and social considerations.

regarding individual workspace control and ownership. Increasingly we work with our virtual systems, whether it be a holiday request online system, video conferencing technologies or a shared file storage repository. The role of organisational goals and culture (as highlighted earlier in Figure 1.1) also play a part here. All individuals within the organisation, and the policies and practices of the organisation itself, are also influenced by financial, technical, legal and social factors.

### ARGUMENT FOR AND AGAINST E/HF

It was implied earlier that it is not always easy to sell the approach of ergonomics and human factors and the aims to jointly improve work and task performance as well as worker and user well-being. We must be prepared for the criticisms or arguments against. Helander (1999) and Pheasant (1986, p. 8) identified some of these, and the authors have directly come across others in their own work. Arguments from engineers, designers and managers can include:

1. 'I can use it' – I have done my user testing on myself, what is the problem? Some software developers will try their own product out and assume that it can be used by all; whilst such expert walkthrough is a very powerful early assessment technique (see Chapter 10), it is almost always not sufficient. It is particularly difficult to 'unlearn' expertise in interacting with systems, and what may seem logical to a designer or engineer who has not



- only designed the ‘user-facing’ elements of a product or system but also been involved in designing the underlying architecture will not automatically be logical to an end-user.
2. ‘We designed it for the average’ – I know we should understand the user, but we were not sure how to do this or even who this was, so we plumped for something in the middle. There are two problems here. First, there is no such thing as an average person; one person might be in a middle range of, say, stature, but generally would not be for leg length, head circumference, shoulder breadth, etc. as well; still less will they be in the middle range of intelligence, strength, vision, hearing, reaction time and many other characteristics as well. Second, where a single characteristic of people is the critical one, and we can identify an average value, it is not usually appropriate to design to it: a doorway for people only of average height, a display that can be read only by people of average eyesight – these are not sensible design choices. Equally, we will not always design for the absolute extremes either – we would not want the cost and inconvenience of houses with door frames 225 cm high nor all instructions in nursery school level language. The ‘user of choice’ is a critical consideration in any design activity and one that will be influenced by many factors, including the nature and business context of the task/product; this must be balanced with accessibility expectations and legislative requirements, ensuring that accommodations of those with additional requirements are either incorporated or feasible with small adjustments.
  3. ‘People are too variable to account for’ – I understand the ideal of accounting for variability across people for a whole set of critical characteristics, but there are too many of these and too much variability. In fact, very early on in any ergonomics contribution to design, we should establish exactly *which* are the critical characteristics that must be accounted for. Moreover, unless there is good reason not to, we will not design for 100% of people but to best suit 95% or 97.5% of the population, whether through establishing the critical end of the range or through adjustability or by fitting to the average in certain circumstances only (see Chapter 27).
  4. ‘People can adapt so why be worried?’ – Why should we bother with all the hassle of ergonomics, people will cope. In this, human beings have been their own worst enemy. We are very adaptable – to poor physical design of workstations, to poor information and instructions, to poor design of jobs. Everywhere, workers and consumers are managing with, or working around, the poor designs of equipment, environments and jobs that they have been given, to keep the company working or to survive in daily life (in fact, when encountering a workplace for the first time, an ergonomist will often glance around for where workers have made adjustments – a label here or an improvised footrest there can be a good clue to where design problems are occurring within the work environment). But such adaptability can come at a long-term cost, to health, satisfaction, spare capacity and reliability if unexpected events or emergencies arise.
  5. ‘It is the workers that make mistakes’ – Let us try to automate everything we can and blame the user if we cannot. Yes, human beings are fallible, they get distracted, tired and forget things. But often those errors are made almost inevitable at some time by the planning, organisation, equipment, job and procedures design and training programmes that have gone before (see Reason’s [1990 and 1997] idea of latent failures and organisation pathogens – see also Chapter 30). Also, to borrow from the legal concept of strict liability, even misuse by individuals is often likely and can often be foreseen by a prudent design team, and should be part of the design decisions that are made. The concept of resilience engineering directly addresses this matter by examining how we can design multiple elements within a work system to anticipate, mitigate and prevent the occurrence and impact of human error. Moreover, it is often the attempts by engineers to design people out of the loop as far as possible that leaves badly designed jobs and circumstances that actually lead to unreliability, by not making the most of the capabilities of people, as Bainbridge (1987) notes in her ironies of automation.



**FIGURE 1.3** The presence of three separate formal and informal sets of instructions (and a threat of removal of entitlement of door use!) suggests that something may be amiss with the design of this set of door controls.

6. 'OK, it is a bad design, but we train them' – Let us increase the adaptability of people to our poor design. Yes, instructions, cues and training are important, but after well-thought-out human-centred design, not instead of it. It is also normally the case that you cannot cure a hardware problem with software changes. Again, sometimes an indication of poor design can be found by the presence of additional instructions (see Figure 1.3) but, where possible, good design should take advantage of pre-existing user skills and experience, and provide intuitive cues and affordances, to encourage correct behaviour (see Norman, 1988 for an extensive discussion of the role of design in encouraging appropriate user interaction).
7. 'People should just do as they are told'. – We have set out perfectly clear instructions or processes; why do not people follow them? Again, we can find a clue here – maybe those processes do not fit into users' expectations of normal behaviour, (if an individual in a maintenance crew is expected to pass all communication via the foreman, but the foreman is 100 m away, and the person who they need to help is right next to them, why should they make the request via the foreman?) or may seem cumbersome or time consuming (e.g. there is much discussion about adherence to and engagement with the WHO surgical checklists that require a pre-operation discussion amongst all the surgical team (Sparks et al., 2013). A better approach is to think *why* are people not adhering to instructions or rules, and how can we ensure safer, more efficient behaviour through a holistic design approach?
8. 'Human factors costs too much'. This is an old argument, and efforts are frequently made to apply cost–benefit reasoning to ergonomics (see Chapter 35). Yes, a contribution to take account of human factors will increase the cost of any design project, implementation scheme or operational overhead – in staff, equipment, study costs and any extra time needed in the system life cycle. But, as we saw from the view on aims of ergonomics earlier, these costs will be outweighed by the saving made on the costs of not getting it right and also the financial and less tangible benefits which will accrue (see Table 1.2).

**TABLE 1.2**  
**Potential Costs and Benefits of Ergonomics/Human Factors**

| Potential Costs of E/HF                            | Potential Costs if No/Poor E/HF     | Potential Benefits from E/HF                     |
|----------------------------------------------------|-------------------------------------|--------------------------------------------------|
| Personnel resources                                | Accidents, incidents and disruption | Improved workforce/customer health and safety    |
| Administrative and contracting resources           | Design retrofit, alteration         | Improved workforce performance                   |
| Technical and equipment resources                  | Product withdrawal                  | Improved workforce satisfaction                  |
| Cost of analysis, design and evaluation process    | Loss and dissatisfaction staff      | Improved client satisfaction and perceived value |
| Capital costs of redesign features                 | Criminal prosecutions               | More effective systems                           |
| Increased system development, design and test time | Compensation claims                 | Legal compliance improved                        |
| Disruption to normal activities                    | Insurance premiums increase         | Better use of resources and capacity             |
|                                                    | Absenteeism, labour turnover, cover | Better labour relations                          |
|                                                    | Recruitment and retraining          | Positive company image                           |
|                                                    | Lower product sales                 | Repeat sales                                     |
|                                                    | Poor company reputation             | Market differentiation                           |

9. ‘Ergonomics is all common sense’. The answer here is yes and no. Some is common sense – but which often seems in short supply without a structured framework within which to identify what needs to be known, apply it and deal with trade-offs (see human factors integration plans – Chapter 9). On the other hand, much of ergonomics and human factors is clearly not common sense; some research findings or guidance may even be counter-intuitive (e.g. the negative impact of underload on attention and task resumability). It is use of the approaches and application of the methods described in this book that will enable us to address the unusual and opaque issues and to rationally bring together into a coherent process and plan the more obvious and straightforward issues.

## CONTEXT FOR APPLICATION OF E/HF METHODS

Since E/HF must have relevance for real settings, the understanding of findings and their application in practice requires a good grasp of context. This same context will also affect society’s view of ergonomics and its value and place in the modern world. It also impacts strongly on choice and application of methods. John Wilson identified the following six features that are particularly important to consider in systems E/HF in a paper written in early 2013 (published as Wilson, 2014) that provide an overview of the context for application of E/HF methods:

1. We need to adopt a *systems focus*, making sure that we move beyond the ‘person-technology’ dyad and considering human interaction within the context of man-made and naturally present systems. One way of making this more manageable is to consider a ‘system of systems’ where we consider an overall system (the example of a hospital is given in Wilson, 2014) encompassing a series of sub-systems, such as beds, ambulances or equipment sets. When selecting appropriate methods, we need to clearly understand the breadth of system that we are considering, and the range of elements within that system that will interact with the designs, workplaces or products that we are specifically examining.
2. We need to remember that behaviour and performance happens in a *context* (see Moray, 1994; Waterson, 2009). This has implications for our selection of methods; if we wish to examine the effect of context, we must capture context and thus, wherever possible,

examine behaviour *in situ*. However, there is still a role for controlled studies in a ‘laboratory’ type context if we wish to capture detailed measures of performance under circumstances which we wish to manipulate (e.g. changing the level of demand in a task simulated to influence perceived workload). What is important is the extent to which we generalise or transfer our results, both to underlying theories, and to real-world contexts.\*

3. A system comprises a number of *interacting parts*. These could be ‘human, technical, information, social, political, economic and organisational components’ (Wilson, 2014). Therefore, our methods need to move beyond examination of linear patterns between discrete elements, to capture complex relationships between people, technologies, artefacts and context. Of particular importance here is the appreciation of not only the range of users a particular product or system might have, but the needs of different stakeholders within that system – in a novel data recording system in a manufacturing context, for example, the machine operator may require data in one form; the line manager may require different data presented and captured over a longer time frame and a production manager may need data from different lines, from different products and machines, in a comparable form.
4. Systems should be considered in a *holistic* manner. This is above and beyond simply considering the context, but acknowledging that people do not work in a social vacuum. Whilst our methods may align to the categories of E/HF outlined earlier, such as physical, cognitive or organisational, we must appreciate that the impact of ‘people on the performance of the human–machine system... and impacts of system design on the well-being of all stakeholders’ (Wilson, 2014) is captured in our methods as needed.
5. Once systems are implemented in a working context, different *emergent properties* will develop. These can be a positive development – for example text messaging was not originally included as a ‘killer app’ for mobile phones in the early twenty-first century (145 billion texts were sent in 2013), (*The Guardian*, 14 January 2014) but, once restrictions on sending texts between different mobile networks were lifted, rapidly it became the most prevalent interaction of users with mobile phones. Conversely, unanticipated behaviour of users can cause people within systems to behave in a different manner to that which had been anticipated by the system designers (as seen in many failed implementations of national IT systems). Therefore, user requirements methods should try to anticipate as much of this emergent behaviour as possible, and capture/manage this behaviour when it does occur.
6. Finally, the place that E/HF has within an organisation has a large influence on the success of *embedding* ergonomics practice, and ultimately influencing change according to E/HF needs. Different organisations place their E/HF expertise (if they have it) in different departments/sections, and some prefer to only use external collaborators to provide an ‘ergonomics view’. This may influence the way in which we design our methods; methods may be more accepted if appropriate and familiar terminology is used, for example. Our own extensive experience in the rail industry has allowed us to become familiar not only with the necessary acronyms and work procedures, but also to appreciate the appropriate levels of complexities and acceptability of our methods in different situations; for example if we wish to conduct some in-depth observation of a night maintenance operation, or a complex signalling task, it is important that our presence is accepted and not seen as an inconvenience, and that we understand enough about the work environment that we

---

\* John and I had many lively discussions about the role of laboratory studies in E/HF. In his 2014 paper, John states ‘laboratory research has its place, but not a primary one’. I would not state this so strongly, and in a footnote in his paper, he notes ‘I do recognise that for some this may be a little extreme, and must point out to the reader that this may not be a universal view in E/HF!’ (SS).

can interpret it without needing to interrupt the operators and ask for any explanations of terms. We joke that once we have been told where the snack drawer is in a workplace, we know that our presence is accepted; achieving this non-intrusive access, whilst maintaining an external analysis perspective is extremely valuable if we are to capture E/HF factors in a systems context.

## ROLE OF E/HF METHODS

Although we looked at definitions of ergonomics and human factors earlier in this chapter, it is more important that E/HF should be seen as an approach (or as a philosophy) of taking account of people in the way we design and organise, in other words, as designing for people. In this view, E/HF itself is primarily a process, to an extent a meta-method, which makes the clear understanding and correct utilisation of individual methods and techniques even more important. This supports the need for this book.

Back in 1986, the NRC Committee on Human Factors reported the development of applied methods to be one of the major research needs for human factors (National Research Council, 1986), and nothing has occurred in the years since to change this. Ergonomics is a science, craft, discipline, practice and a technology and thus has need of techniques for both data collection (basic or functional data), analysis and application. The debts we owe to other disciplines are obvious. However, as we have gained experience and confidence and as our armoury of knowledge and methodology have grown, so the debt to other disciplines has been, and is being, repaid. Methods developed or adapted within E/HF are employed by psychologists, engineers, computer scientists, management scientists or health care professionals in turn, just as we are constantly enlarging our own human performance database through results of human-machine systems evaluation.

Bearing in mind the applied nature of ergonomics, we have both general methods and approaches (e.g. observation, subjective assessment, data logging and experiments) that can be applied at many stages of the design process, and then specific tools that are used when we have specific types of input to the design process that we wish to make. These can be classified into five groups: data about people, systems analysis and development, evaluation of system performance, assessment of effects on people and the organisation of E/HF management programmes.

## METHODS FOR THE COLLECTION OF DATA ABOUT PEOPLE AND THEIR NEEDS

Our first methodological need is for data about people and this can cover all characteristics: physical size and strength, physiological capability, sensory characteristics, mental capacities and psychological responses. In addition to collection of data about these human characteristics, it is also important that we establish users' motivations and desires – this is particularly valuable when we are trying to balance priorities of, for example usability and aesthetics. Just as important as the collection and reporting of data is the generation of design and evaluation criteria from these. For example given data on the population range of arm reaches in different directions, what advice can be given on placement of frequently used rotary controls? Or again, given data on working memory limitations, can these be adapted to form design guidance on numbers of different codes that should be used in a coding system? Methods used to produce data about people comprise much of the scientific base of E/HF.

## METHODS USED IN SYSTEMS ANALYSIS AND DESIGN

The second input of E/HF is its contribution to the analysis and design process. This relates specifically to methods that assist in the analysis and design stages of development of equipment, workplaces, technologies, jobs, work processes or buildings. In essence, we need methods to *analyse* current or proposed systems (analysis strictly meaning to resolve the system into its constituent

elements and critically examine these) and then to *synthesise* data (i.e. build up a coherent whole by putting elements back together) into ergonomically sound concepts, prototypes and final designs (see Chapter 9). Specifications produced out of this process must be capable of transference into design requirements, criteria and data, and also have reasoned justifications, in order that ergonomists can work sensibly with engineers and designers.

## **METHODS TO EVALUATE HUMAN–MACHINE SYSTEM PERFORMANCE**

In part, analysis at the start of development may involve evaluations of an existing system's performance. Certainly, we must evaluate system performance during and at the end of development to support cost–benefit analysis and enable benchmarking of system performance. Many measures can be defined for this, and one challenge today is the search for measures of system performance that extend beyond the typical direct (but often sterile) measures of times and errors other than. Manufacturing system performance, for instance, may be assessed by means of production output rates and product quality levels, but we could also use machine utilisation rates, minimisation of finished stocks or work-in-progress, raw material wastage, speed of response to changed schedules, accident rates, sickness or other absence, and job attitudes and job satisfaction measures. Similarly, although a computer interface can be assessed in terms of time taken and errors made in performing a sequence of tasks, more interesting measures might be 'extent of system explored', 'willingness to change direction', 'quality of finished work', etc. (Johnson and Wilson, 1988). New technologies, such as detailed data recording and sensing equipment, along with the increased capacities and reduced intrusion of video tools, make detailed longitudinal data collection of complex interactions much more feasible than in the past, and can support these more detailed and valid approaches to analysis of system performance.

Any evaluation of subsequent system performance where there is a human factors input into system development is also, in an interesting closing of the loop, an evaluation of just how well E/HF was applied to the design.

## **METHODS TO ASSESS DEMANDS AND EFFECTS ON PEOPLE**

As we have seen, E/HF has twin aims in its contribution to design and development: improvements for the job-holder or user *and* improvements for the producer, employer or organisation. As a result, any system assessment should be carried out in terms of the demands made on people and of effects on their well-being, as well as accounting for system performance. There is an argument that the demands made on people should in fact be viewed as an implicit part of system performance and as such, their assessment should not be seen as separate; however, on balance, it is more persuasive that such a distinction will emphasise the twin, yet interdependent, aims of E/HF. Many methods can be applied to assess the effects that different environments, jobs or equipment have on people. Such impacts might be medical, physical or psychological in nature, and methods will vary from recording of measureable phenomena (e.g. heart rate) to observation of people's affective states (e.g. boredom). Again, technological developments have dramatically changed the extent to which such recording methods are feasible to implement in a workplace or complex simulation – for example it is now possible to capture a person's expression and automatically infer their emotional state (e.g. el Kaliouby and Robinson, 2005). In almost all circumstances however, the data collected are not useful by themselves but must be interpreted and any effects for the overall system design inferred by looking at the context of the effect and the interacting factors that may have contributed to that demand or effect; this is again a large part of the ergonomist's input. Moreover, if assessment methodology is developed appropriately, then data obtained can be generalised to become part of our first input, the basic data on people.



## METHODS TO DEVELOP E/HF MANAGEMENT PROGRAMMES

Our fifth input concerns the management of E/HF programmes. Methods are required here for two situations, although there is not a clear distinction between them. Firstly, there are ergonomists who are working within companies, and they are often doing this in very small groups or even on their own. Secondly, we have what is termed ‘devolving’ (Wilson, 1994) or ‘giving away’ (Corlett, 1991) ergonomics expertise. It is unrealistic to expect all enterprises to employ only ergonomists (either as employees or outside consultants) to handle their ergonomics; it is inappropriate also in many circumstances. In many areas – for instance health and safety, job redesign, workplace layout – the ergonomics profession must provide methods which allow the development of appropriate strategies and which support the management of programmes which can be run as a part of normal company activities. Nobody is advocating that untrained staff handle all human factors in a company. However, design engineering, systems engineering, interaction designers, health and safety advisors, line management and production workers can all make considerable contributions; the methods and support that ergonomists give to them must include enabling them to recognise when specialists must be brought in. This fifth input is certainly the ‘messiest’ area of E/HF methodology, and it embraces aspects of all other inputs as well as participatory ergonomics, systems implementation and so on (see Chapters 28, 32 and 34). Nonetheless, it is an increasingly important area as E/HF moves more and more into real companies with real problems as well as maintaining its position within research laboratories and universities.

## CLASSIFICATION OF METHODS

Classification of ergonomics methods in terms of the parts or stages of any models of E/HF is difficult. Meister (1985) attempts a gross distinction and divides his behavioural methods into the analytic techniques employed during the development of systems and the measurement methods employed to evaluate functioning systems. He does, though, recognise overlaps, particularly that many measurement methods are also used during system development. Similarly, at a first level, Sharples and Megaw, in Chapter 18 of this book, distinguish analytical methods or measures – broadly those based on theory, modelling and prediction – from empirical ones – those based on observation or experiments with an actual situation.

The following tables are adapted from one which was published in the earlier editions of this book, and provides an overview of the different types of methods that are needed to address all aspects of a work or interactive system. Six core types of methods are identified, represented within each of the sub-tables. The six methods types are:

- a. *General methods*: Representing approaches of data collection and/or analysis that can be used with a range of different design or evaluation goals, across a work system, and normally within any of the method categories (b to f).
- b. *Collection of information about people, equipment and environment*: Typically ‘baseline’ and descriptive measures of either an individual or environment’s state or characteristics. For an individual, this could be an inherent state, such as their anthropometry or age, or one that is transient and affected by external factors, such as their attitude; similarly for an environment, this could be a core characteristic, such as the size of a room, or its current thermal state.
- c. *Analysis and design*: Tools that move beyond description to provide links between the data collected and underlying influences, actively supporting the design process.
- d. *Evaluation of user and system performance*: This will often use methods of type (b) as a baseline to understand both the performance outputs of a system, and the factors that may influence those outputs.

- e. *Evaluation of demands on and effects experienced by people*: Capturing data from people about their physiological and psychological responses to systems, experiences and tasks.
- f. *Management and implementation of ergonomics*: Toolkits to support the integration of E/HF, normally focussing on early consideration of E/HF requirements in work/system design.

The tables are not intended to be exhaustive, and categories are not mutually exclusive, but aim to provide an indication of the types of methods that an ergonomist should consider using, and their outputs. One notable difference between the contents of these tables compared to the similar table in third edition of this book is the emergence of complex modelling and technology-enabled data collection approaches, which have become both technically feasible, cost-effective and of reduced intrusion compared with a decade earlier (Table 1.3).

## DEBATES AND DISCUSSIONS AROUND METHODOLOGICAL APPROACHES

As far as methods are concerned, according to an old expression, the proof of the pudding is in the eating. A method which to one researcher or practitioner is an invaluable aid to all their work may to another be vague or insubstantial in concept, difficult to use and variable in its outcomes; of course, individual E/HF practitioners and teams develop expertise in and familiarity with particular tools. More than this, the validity, reliability and sensitivity of methods will be application specific. We may need examination of utility and generalisability also in order to select or prioritise between methods. However, a number of traditional ‘debates’ persist within the discipline, and we consider some of them here.

Once all possible measures are considered, we begin to see some of the problems of classification and selection. Kantowitz (1992) lists 46 possible indicators of nuclear power plant safety, split into seven categories (e.g. operations, quality programmes, management/administration) and reports that ‘no single indicator was by itself an adequate measure... [nor were any] optimal for predicting plant safety’ (p. 391).

It is not surprising therefore that we will often look to use more than one, and often several, methods and measures in any one study. Indeed, as E/HF has a bounty of techniques and methods to draw upon, the issue should not be one of A versus B but of selecting a subset of our broad range of available methods that will be appropriate to the issue at hand.

This is particularly so when we are carrying out evaluations in the field. Technically, if this is done formally, it is known as triangulation (see Denzin, 1970 or Webb et al., 1972) (and, despite the name, requires only two or more methods or approaches to be applied independently). Guion et al. (2011) identify five types of triangulation: Data (using different source of information); Investigator (using different people in the analysis process); Theory (using multiple perspective to interpret data); Environmental (Using different location settings or contexts to capture data); and the most commonly considered type of triangulation, Methodological Triangulation. Methodological Triangulation refers to the independent use of multiple distinct qualitative or quantitative methods to address the same research question. The application of triangulation will improve the effectiveness of a study or our confidence in any findings; weaknesses in one method can be balanced by strengths in another. To take one simple example, only by questioning and observing operators in complex systems *and* also recording their concurrent or retrospective verbal reports for subsequent protocol analysis can we begin to understand something about their decision making activities. A multiple-methods study may utilise a mixture of qualitative and quantitative techniques, in field and laboratory settings. Such an approach in the field is often known as contextual inquiry.

Only by use of several of these methods may a full evaluation be possible in any one situation, and thereby effective suggestions be made for redesigning job content, tasks, workstations and environments. However, a word of caution is in order. It is all too easy to fall into the trap, once an investigation has started, of measuring everything possible ‘just in case’! This can lead to results



**TABLE 1.3**  
**Classification of Methods, Techniques and Measures Used with Ergonomics/Human Factors**

| Method                                                                       |                                                                                                                                                                                         |                                                                                                                           |
|------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| Group                                                                        | Example Techniques                                                                                                                                                                      | Typical Measure/Outcome                                                                                                   |
| <b>a. General methods</b>                                                    |                                                                                                                                                                                         |                                                                                                                           |
| Direct observation                                                           | Human recording: observation checklists, expert rating, photographs, diagrams                                                                                                           | Event frequency; sequence, strategy analysis; performance time, error, accuracy; participant behaviour                    |
|                                                                              | Technology-led recording: videos, audio, photos, data and event logging, location tracking                                                                                              | Interaction patterns; link/sequence analysis; emotions                                                                    |
|                                                                              | Performance measures                                                                                                                                                                    | Time, error; strategy descriptions                                                                                        |
| Indirect observation                                                         | Archival analysis, artefact analysis, data mining                                                                                                                                       | Event counts; covariation analysis; critical incident analysis; temporal sequence analysis; absenteeism, sickness records |
| Participant reports                                                          | Verbal protocol, surveys, questionnaires, rating, ranking, scaling, diaries, checklists                                                                                                 | Attitudes; perceived effort; preferences                                                                                  |
| Practitioner–participant communications                                      | Group discussions, interviews, participant-led observations                                                                                                                             | Themes; thoughts and preferences; ideas                                                                                   |
| Work system instrumentation                                                  | Screen capture, eye tracking, voice recording, data logs                                                                                                                                | Scenario description; eye movement analysis                                                                               |
| Experiments                                                                  | Inferential and correlational designs, multiple and individual independent variables, analysis of covariables                                                                           | Causative analysis; likelihood of relationships compared to chance; statistical trends and patterns                       |
| Literature and data interpretation                                           | Description, statistical analysis                                                                                                                                                       | Patterns and prevalence; links, networks                                                                                  |
| Standards and recommendations                                                | Standard adherence check (e.g. BSI [British Standards Institute], ISO [International Organization for Standardization])                                                                 | Official or unofficial quality statement                                                                                  |
| Prediction and modelling                                                     | Analytical hierarchical process, human reliability assessment, simulation, cognitive work analysis                                                                                      | Predicted behaviours and probabilities                                                                                    |
| Multiple methods                                                             | Triangulation, mixed methods                                                                                                                                                            | Validity and reliability indications                                                                                      |
| <b>b. Collection of Information about people, equipment and environments</b> |                                                                                                                                                                                         |                                                                                                                           |
| Physical measurement                                                         | Anthropometry, 3D scanning, ultrasound, fitting trials, photographs                                                                                                                     | Dimensions, percentiles, population descriptions                                                                          |
| Physiological measurement                                                    | Biomechanics, goniometry, stadiometers                                                                                                                                                  | Spinal shrinkage, angles, strength                                                                                        |
|                                                                              | Sway measurement, posture and gait, heart rate variability, muscle contraction, brain activity (e.g. electroencephalography, functional near-infrared spectroscopy, magnetic resonance) | Brain activity, posture, gait, movements                                                                                  |
| Environmental measurement                                                    | Hygrometer, thermometer, sensors, light/sound meter, accelerometer                                                                                                                      | Humidity, temperature, vibration, light/sound levels                                                                      |
| Perceptual/cognitive assessment                                              | Visual acuity, colour blindness, stereopsis.                                                                                                                                            | Perceptual ability                                                                                                        |
|                                                                              | Psychophysics (method of limits, constant stimuli, adjustment)                                                                                                                          | Thresholds and levels of perception, sensitivity                                                                          |
|                                                                              | Mental or cognitive tests (e.g. intelligence, spatial ability), response times, skill tests                                                                                             | Predictions of performance, individual characteristics                                                                    |

(Continued)

**TABLE 1.3 (Continued)****Classification of Methods, Techniques and Measures Used with Ergonomics/Human Factors**

| <b>Method</b>                                    |                                                                                                                                                     |                                                                                               |
|--------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| <b>Group</b>                                     | <b>Example Techniques</b>                                                                                                                           | <b>Typical Measure/Outcome</b>                                                                |
| Social and organisational measures               | Cultural probes, photo diaries, historical analysis, network analysis, social media analysis, sentiment analysis                                    | Communication patterns, habits, interactions, networks, feelings, team behaviours             |
| Knowledge                                        | Written, audio, video records, interviews, protocol analysis, conceptual mapping, goal decomposition                                                | Rules, practices, reasoning, explanations                                                     |
| Models                                           | Artificial intelligence models, biomechanical models, logical mathematical models                                                                   | Processes, sequences, limits and capabilities                                                 |
| Task analysis                                    | Hierarchical, tabular, link analysis                                                                                                                | Task completion methods, action sequences                                                     |
| <b>c. Methods to support analysis and design</b> |                                                                                                                                                     |                                                                                               |
| Task analysis                                    | Hierarchical, tabular task analysis, requirements analysis, cognitive task analysis, task action grammars, operation sequence diagrams, flow charts | Requirements for people and systems, data to support prediction of consequence of task change |
| Expert analysis                                  | Guidelines, heuristics, standards, cognitive walkthrough, Delphi technique, likelihood matrix                                                       | Weightings and priorities, qualitative reports, critical issues, risk analysis                |
| Introspection/protocol analysis                  | Repertory grid, concurrent/retrospective verbal protocol, diaries, shadowing                                                                        | Behaviour transitions, rules, strategies, expertise, knowledge models                         |
| User models                                      | Mental models, task action grammar, GOMS (Goals, Operations, Methods, Selection)                                                                    | Strategies, beliefs, priorities, goals                                                        |
| Statistical analysis                             | Signal detection theory, information theory, reliability assessment                                                                                 | Performance measures, likelihood of outcomes                                                  |
| Models                                           | Rich pictures, soft systems methodology                                                                                                             | Performance predictions                                                                       |
| Simulation                                       | Mathematical and computer models, virtual reality, computer-aided design, personas                                                                  | System capacity and capability                                                                |
| Method study                                     | Link analysis, layout analysis, process, flow, time charts; micromotion                                                                             | Movements, times, actions, frequencies                                                        |
| Work measurement                                 | Time study, activity analysis, synthetic analysis                                                                                                   | Times, standards, task sequence, simultaneity, frequency, demand                              |
| System-level representation/modelling            | Influence diagrams, STAMP (System-Theoretic Accident Model and Processes), soft systems methodology, Accimaps, cognitive work analysis              | System interactions, boundaries and emergent behaviours                                       |
| Prototyping                                      | Wireframing, paper prototypes, virtual prototypes                                                                                                   | Design representations, user/stakeholder views on designs                                     |
| Creativity techniques                            | Scenarios/use cases, stop/start animations                                                                                                          | Novel design ideas, representations of user perspectives                                      |
| Participative methods                            | Co-design, drama in design, role playing, design decision groups, co-operative evaluation, living labs                                              | User-informed design suggestions, user requirements                                           |

*(Continued)*

**TABLE 1.3 (Continued)****Classification of Methods, Techniques and Measures Used with Ergonomics/Human Factors**

| Method                                                               |                                                                                                                                                                                   |                                                                                       |
|----------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| Group                                                                | Example Techniques                                                                                                                                                                | Typical Measure/Outcome                                                               |
| <b>d. Evaluation of user and system performance</b>                  |                                                                                                                                                                                   |                                                                                       |
| Work systems analysis                                                | Cognitive work analysis, checklists, walkthrough, expert assessment, critical decision method                                                                                     | Time, reaction time, strategies, processes                                            |
| Usability evaluation                                                 | Observation, protocol analysis, expert analysis, interaction analysis, screen recording                                                                                           | Accuracy, errors, opinions, attitudes, satisfaction, workload                         |
| Work system instrumentation                                          | Environmental measures, frequency of use (e.g. footfall measures), performance logs, eye tracking                                                                                 | Measurement against comparison, baselines, norms; fit to requirements; acceptability  |
| Participant reports                                                  | Scaling, rating, questionnaires                                                                                                                                                   | Comfort, annoyance, acceptance, pleasure                                              |
| Performance measures                                                 | Work rate, waste, errors, communications analysis, behavioural markers, secondary task measures                                                                                   | Measurement against comparison, baselines, norms                                      |
| Modelling and simulation                                             | Critical path analysis, keystroke level model                                                                                                                                     | Strategies, task sequences                                                            |
| Self-recording                                                       | Gripe button, diaries, event recorder, crowd-sourcing                                                                                                                             | Problems, incidents, potential improvements                                           |
| Text analysis                                                        | Readability formulae, judgements, scan/read tests                                                                                                                                 | Normative scores, age ratings                                                         |
| Human reliability analysis                                           | Error modelling, error prediction, error and incident classification                                                                                                              | Error analysis, causal analysis, error probabilities                                  |
| Accident and safety reporting and analysis                           | Event analysis, incident analysis, Accimaps                                                                                                                                       | Incidence, severity, aetiology                                                        |
| <b>e. Evaluation of demands on and effects experienced by people</b> |                                                                                                                                                                                   |                                                                                       |
| Physical response                                                    | Subjective assessment, perceived exertion, sickness ratings, physical changes (e.g. stadiometer), psychomotor performance                                                         | Ratings, performance change, physical measures                                        |
| Psychological response                                               | Subjective response, performance measures, task strategies                                                                                                                        | Experienced load, perceived demand                                                    |
| Posture and activity analysis                                        | Automatic posture analysis (e.g. CODA, computer vision), Postural assessment tools (e.g. RULA [Rapid Upper Limb Assessment], REBA [Rapid Entire Body Assessment]), force platform | Posture descriptions, rating of risk of injury rating, body part angles and positions |
| Physiological measures                                               | Heart rate variability, galvanic skin response, oxygen uptake, brain activity                                                                                                     | Objective comparative data                                                            |
| External demands                                                     | Time pressures, frequency of demands                                                                                                                                              | Temporal analysis of activity patterns                                                |
| Fatigue measures                                                     | Critical flicker frequency, task performance, reported fatigue, blink rate                                                                                                        | Visual fatigue scores                                                                 |
| Environmental response                                               | Sweat rate, body temperature, hearing impairment, contrast sensitivity, sensation loss                                                                                            | Measurements to compare to norm/baseline                                              |
| Stress measures                                                      | Galvanic skin response, subjective report techniques                                                                                                                              | Measures to compare to norm/baseline, subjective ratings                              |
| Job and work attitude measurement                                    | Rating scales, interviews                                                                                                                                                         | Satisfaction, user needs, job characteristics                                         |

*(Continued)*

**TABLE 1.3 (Continued)****Classification of Methods, Techniques and Measures Used with Ergonomics/Human Factors**

| Method                                                           |                                                                                                                              |                                                                                                                 |
|------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| Group                                                            | Example Techniques                                                                                                           | Typical Measure/Outcome                                                                                         |
| <b>f. Management and implementation of ergonomics</b>            |                                                                                                                              |                                                                                                                 |
| Ergonomics management, organisation analysis, project management | Human factors integration, early human factors analysis, lifecycle management, MANPRINT (Manpower and Personnel Integration) | Early consideration of E/HF in system/product design lifecycle, inclusion of E/HF in design/operational process |
| Implementation                                                   | ISO tools, heuristics, embedded E/HF experts in organisation                                                                 | Change to implementation approach                                                                               |
| Cost–benefit analysis                                            | Cost estimates, activity analysis                                                                                            | Financial analysis of impact of E/HF                                                                            |
| Participative methods                                            | Co-design, focus groups, elicitation techniques                                                                              | User influence in system/product design                                                                         |
| Ethics in ergonomics                                             | Guidelines                                                                                                                   | Adherence to guidance                                                                                           |

that are difficult to interpret or, worse still, an analysis phase that uses large amounts of study resources to no clear or useful purpose.

In selecting an appropriate approach and set of methods for any particular need, we will need to weigh up the relative merits of a quantitative or qualitative approach, so-called objective or subjective methods, and studies in the field or laboratory.

### QUALITATIVE/QUANTITATIVE DEBATE FOR METHODOLOGY IN ERGONOMICS/HUMAN FACTORS

Qualitative approaches and methodology are being increasingly used in E/HF and sister disciplines such as organisational management, product design, psychology and engineering, and chapters throughout this book refer to its much increased acceptability and use. Although there has been considerable debate, even rancour, about its place in disciplines such as psychology, there is good reason to think that E/HF will benefit from the explicit inclusion of qualitative methods and tools in teaching, fundamental research and practice. This, in turn, would help to support the consideration of social, emotional and philosophical factors when conducting both theoretical and practical studies (Hignett and Wilson, 2004a,b). E/HF can be located on the cusp of the sciences and the humanities, and in the centre of the qualitative–quantitative continuum; it has much perhaps in common with anthropology, where the unit of analysis is interaction, in contrast to psychology where the unit of analysis is the individual.

Moray (1994, 2000) has recognised that sociologists, anthropologists and ethnographers offer a methodology more sensitive to the context of analysis of work, rather than methodologies which result in generalisations based on quantitative analysis, but that there is need to generalise findings of qualitative study to understand other different systems. There should be no claim made for a general superiority or even a preference for either of the qualitative or quantitative philosophies, approaches, methods and measures. Each should be used as needs and circumstances dictate, and we need to be clear about differences and similarities in scope, approach and outcomes – see Table 1.4 from Hignett and Wilson (2004a).

### ‘OBJECTIVE’ AND ‘SUBJECTIVE’ METHODS

The recent debates over the relative merits of quantitative and qualitative approaches interact with debate over so-called objective and subjective measurement (Wilson and Nichols, 2002). This has been an issue from the earliest days of the ergonomics and human factors discipline, with early

**TABLE 1.4**  
**Dimensions of Qualitative and Quantitative Methodologies**

| Qualitative Dimensions                                                           | Quantitative Dimensions                                                                                   |
|----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| Words, understanding                                                             | Numbers, explanation                                                                                      |
| Purposive sampling, inductive reasoning                                          | Statistical sampling, deductive reasoning                                                                 |
| Social sciences, soft, subjective                                                | Physical sciences, hard, objective                                                                        |
| Practitioner as a human being to gather data, personal                           | Researcher, descriptive, impersonal                                                                       |
| Inquiry from the inside                                                          | Inquiry from the outside                                                                                  |
| Data collection and analysis intertwined                                         | Data collection before analysis                                                                           |
| Creative, acknowledges of extraneous variables as contributing to the phenomenon | Predefined, operationalised concepts stated as hypotheses, empirical measurement and control of variables |
| Meanings of behaviours, broad and inclusive focus                                | Cause and effect relationship                                                                             |
| Discovery, gaining knowledge, understanding actions                              | Theory/explanation testing and development                                                                |

Source: Adapted from Hignett, S. and Wilson, J.R., *Theor. Issues Ergon. Sci.*, 5, 517, 2004b.

views or implications that if subjective measurements do not match objective measurements, then it is the former which are biased (e.g. Poulton, 1975).

However, what do we mean by objective measurement? Different forms of apparently objective measurement may show many elements of possible observer bias and of subjectivity in the actual selection of measures, criterion levels and methods. This was illustrated for us in work on mental model identification and representation; the form of mental models reported in the literature often seemed dependent upon the scientists' choices of experimental and measurement method, in turn based on their preconceptions about the form of mental models (Rutherford and Wilson, 1992).

Ergonomics and human factors deal with many concepts which are difficult to define explicitly and therefore have no clear agreement on method of measurement. We work with the complexity of human beings, individually and as social groups. In cases of mental workload, situation awareness, fatigue, mental models and presence for instance, it is quite possible that what we are measuring with objective (performance or physiological) methods may be different – if related – entities to what is measured with subjective methods (participant ratings, interviews etc.). This means that such concepts should be clearly, often operationally, defined in relation to the measurement method being used, rather than independently of it. Kanis (2014) contributes to this debate in his discussion of validity and reliability in E/HF tools, highlighting the perils of 'measurement by fiat' (the imposition of numerical values by arbitrary definition after presuming a relationship between obtainable observations and the property of the subject under investigation [Torgerson, 1958]). We would argue that subjective methods are extremely valuable firstly as tools to obtain information about a task or system quickly, with minimal intrusion, and without the need for specialist equipment, but secondly, and probably more importantly, to capture the user's perception about the system, as this perception may not only affect the user's attitude, but also their selected behaviours and interactions (see Chapter 18 by Sharples and Megaw). Yes, this opinion may be biased in some way (and of course, we should use techniques of scale design and good practice to ensure that this bias is minimised [see Chapter 4]), but this 'biased' opinion has a direct influence on the user's interactions and behaviours, and is important to capture.

Indeed, the word 'subjective' is a challenging one, and is often associated with an implied derogatory tone, assuming that in some way, a method that is subjective is inherently inaccurate. This is not surprising when we consider two alternative definitions of subjective from the Oxford English Dictionary: The first definition, which might be paraphrased as 'of the subject (i.e. participant)' is *Relating to a person who is subject; belonging to or characteristic of a political subject; submissive, obedient.*, yet a further definition is *Of, relating to, or proceeding from an individual's thoughts,*

*views, etc.; derived from or expressing a person's individuality or idiosyncrasy; not impartial or literal; personal, individual. OR Existing in the mind only, without anything real to correspond to it; illusory, fanciful.* (Oxford English Dictionary, 2014).

The first definition simply uses the adjective to denote the source of information (the 'subject'), whereas the second places the notion of subjective data in opposition to that collected through 'objective' means. In fact, much of our skill in development of methods allows us to collect 'subjective' data in an 'objective' manner (see Zhang et al. (1996), Jian et al. (2000) and Pickup et al. (2005) for examples of such approaches). Annett (2002a) in a target paper, and several commentary papers in the same issue of ergonomics, covers very well the topic of objectivity and subjectivity in ergonomics measurement and assessment.

## FIELD AND LABORATORY STUDY

Within the sporadic debates about the nature of ergonomics and human factors over the past decade, there has been much discussion of the merits of field and laboratory study, contrasting formal and informal methodological approaches and setting the advantages of control against those of veridicity (see Wilson, 2000).<sup>\*</sup> Moray (1994) believes that ergonomics only makes sense in the full richness of the social settings in which people work, and Wilson (2014) returns to this theme. The influence of contextual factors – such as interactions between individuals, the formation and relationships of teams, individual motivations, computer-based support systems and organisational structures – on work in practice must be understood before we can decide on techniques for the measurement and prediction of the outcomes of people's work, or on recommendations for the design of systems. Our question is therefore how do we capture this contextual influence, but retain the power that can be obtained from the structure and control that we can have the opportunity to impose in a laboratory context.

Before we consider the relative merits of the laboratory and field, it is important that we distinguish the laboratory from 'experiments'. The experimental paradigm, the description of elements as independent or dependent variables, and the structure imposed on measurement, which may then lead to use of statistical analysis in some form, are useful tools that underpin scientific research (see Chapter 2). But this paradigm does not need to be restricted to a controlled laboratory context. For example if we are collecting data in a naturalistic driving study (such as was done in the 100 car study, Neale et al., 2005) we may specify one influencing ('independent') variable as being driver experience, and examine the number of driving errors or incidents recorded for drivers with varying numbers of years of driving. This allows us to utilise not only the experimental paradigm and infer causality (the influence of experience on driving performance) but *also* capture data in context. So the use of an experimental paradigm does *not* automatically imply the collection of data in a tightly controlled laboratory environment.

Laboratory studies have long been the dominant approach in, for instance, physical ergonomics research. This is a vital source of information and insight about isolated work variables, but may not be the most valid approach to understand work in practice and certainly will not replace the need for field study. The very nature of the traditional controlled laboratory environment means that the complexities and uncertainties of work environments are, by definition, not being simulated.

Sometimes, however, a field context simply does not yield the appropriate types of data, in an appropriate form, to enable us to study the research question that we wish to study. The context may be safety critical, or the methods which we wish to apply (e.g. concurrent verbal protocol, or frequent subjective ratings) may be intrusive and change the nature of task completion. If, for example, we wished to implement a new interface for representation of text-based communications in ATC,

---

<sup>\*</sup> As mentioned earlier, this was an area about which John and I had much discussion. This section was written after John's death, and I fear he would disagree with some of the points here (or at least would have argued that they should be stated differently), so for a pure 'Wilson' view, the reader is referred to the third edition of this book. (SS).

it would be sensible to use a laboratory study to examine some of the basic issues associated with screen layout or complexity before implementing the tool in an active control context. A contribution that the field of E/HF has to make to the conduct of laboratory studies is in establishing approaches that allow us to *transfer* between the lab and the field; ensuring that we fully understand the extent to which we can generalise the data we collect in a laboratory context (so ensuring we report clearly what was and was not controlled for example) and designing those laboratory contexts and tasks to an appropriate level of complexity that allows the difference between behaviour in a field and laboratory context to be as close as possible. Simulation technologies are now extending to not only be able to capture aspects of individual control elements of work, but also the communications and some contextual elements (Sharples et al., 2011).

It is also our responsibility to ensure that any laboratory studies that we do design manipulate and examine the appropriate variables that are of importance in a field context. It is then also important to return to the field, to appreciate the roles of extraneous and influencing variables (e.g. environmental disturbances, work flow, individual motivation, inter-personal relationships and team relationships (Wilson et al., 2003) that have been controlled in a laboratory setting.

Work is difficult to study in practice, but ‘when the unit of analysis is interactions, then field research is arguably the main methodological approach for ergonomics’ (de Keyser, 1992). ‘Methodologically explicit field research is vital for the core purpose of ergonomics, investigating and improving interactions between people and the world around them’ (Wilson, 2000, p. 563).

We must be aware that field study naturally includes context from the environment, supervision, motivation and disturbances, but the very presence of the investigator will create an additional context which, without care, can invalidate the study. Well-planned, sensitive and long-term studies can minimise any influences (Webb et al., 1972; Bernard, 1995; Hammersley and Atkinson, 1995; Denzin and Lincoln, 1998). Ethnomethodologists talk of the ‘participant observer’, in special cases when the observer becomes a member of the group being studied and also generally, because they also participate as an actor in the workplace by their very presence as a researcher.

Our argument is for careful consideration of the use of field and laboratory contexts, and understanding of the value and contribution of both. The worst-case scenario is a field study that captures the wrong or insufficient data to allow detailed understanding of the research question, or a laboratory study that either fails to control variables, or is so complex that it is not possible to infer the influence of a useful isolated independent variable. If time and resources allow, an excellent approach can be seen as: starting in the field to identify the problem statement; collecting data using a mixed methods approach, using the field or laboratory as the question or researcher expertise demands and returning to the field to reflect upon the impact or context of findings. This is certainly not an argument against laboratory research – it is important to acknowledge that both laboratory and field studies are necessary – but is an argument for a better balance. Nor is there usually a stark either/or choice to be made.

## QUALITY OF METHODS AND MEASURES AND THEIR CHOICE

In a recent commentary at the end of a special issue on ergonomics models and measurement, Annett (2002b) gently reminded us that ergonomics texts (the second edition of this one included!) do not always treat issues of the quality of ergonomics methods to the extent needed. In particular, he suggests that we have not properly considered the reliability and validity of our methods and measurement. Kanis (2014) with responses by Hignett (2014), Stanton (2014) and Vink (2014) continues this discussion. In fact, there are a number of ways in which we might judge the adequacy and quality of methods and measures, and therefore, we have a number of selection criteria for methods. These are discussed in the following text and the reader should see also any good research methods texts, such as Dane (1990), Frankfort-Nachmias and Nachmias (1996) or Robson (2011). It should be pointed out that we will rarely, if ever, be able to satisfy all of these criteria but equally we will rarely if ever need to for any particular study or investigation.



## VALIDITY

The concept of validity can be defined as whether something measures what it claims to measure. This encompasses both whether a method looks like it measures what it actually measures, and demonstrates not only the importance of using carefully developed methods but also being very clear about what a method actually does or does not measure. E/HF presents a particular challenge here, as many of the terms we use, like workload, attention or teamwork, may have colloquial meanings that are understandable to the general public, but also specific meanings within the context of E/HF science. Therefore, one of our responsibilities, as referred to in the earlier discussion about devolving ergonomics, is to ensure that we clearly communicate the capabilities and appropriate implementation context of any method that we develop.

There are many different forms of validity; some types are embedded within others, some are very similar but are given quite different names.

*Construct validity* represents the extent to which a measure represents the concepts that it should, and particularly that it does not represent those that it should not. *Content validity* is complementary to construct validity and asks whether a measure is complete and measures all relevant parts of the concept it claims to measure.

Within construct and content validity, we might look for: *convergent validity* (the measure is consistent with outcomes from two or more different measures – a form of triangulation); *divergent validity* (the measure does not correlate closely with measures of different concepts); and *face validity* – see the following text.

*Internal validity* is the extent to which a measure allows a plausible demonstration of causal relationships (i.e. the identified cause does in fact lead to the assumed effect) and *external validity* relates to the generalisability of results or findings to other situations, times or contexts.

*Criterion validity* – the extent to which a measure is able to predict an outcome for one variable, based on data from another variable, has two forms: *concurrent validity* – whether the new measure compares well to outcomes of an existing ‘accepted’ measure (a limited form of and related to convergent validity) and *predictive validity* – how well the measure predicts what we eventually see happen in the real world.

Related to criterion validity, but important when transferring findings from a simulated or laboratory context to the real world, is *behavioural validity*. This describes whether behaviour observed or recorded in one context is consistent with that which would be found in a real-world context.

Finally, *face validity* is, literally, the validity of a measure or method ‘on the face of it’. Sometimes seen as a form of construct validity, face validity considers the degree of consensus that a measure actually represents a particular concept and is strongly influenced by the look or appearance of a measure. It can also denote how acceptable a measure or method is to a particular set of stakeholders, which can be an important consideration when implementing measures in a real working context.

## RELIABILITY

It is sometimes said that reliability is a special sort of validity, and certainly, it is required in order to be sure of validity. Reliability comes in many different forms and manner of assessment as well, but essentially asks whether we would get the same results and interpretations if we repeatedly use a method or measure; it is directly related to the amount of bias or error we may have allowed to creep in. Bias and error may be from the investigator (and her arrangements) or the study participant, and can have conscious or sub-conscious origins. We may look for assessment of reliability through inter-rater comparisons, use of test/retest procedures, application of different forms of measure at different times or on different groups, or through split half techniques whereby the measure is somehow divided into two for each participant (e.g. divide the questions in a scale into two halves), and compare the scores within participants.



## GENERALISABILITY

Literally this refers to how well a method, measure or measurement, and data from that method, will generalise into other domains, situations, settings or populations. Influences and limitations on generalisability of methods may include the use of domain-specific terminology (e.g. in a specially designed questionnaire for health professionals) or the skills and experience of the participants being studied with respect to those to whom the results will be generalised. Generalisability of data is affected by the influence of context-specific variables on the behaviours or activities of interest in both the original domain and the domain to which the data is being applied. This can be difficult to determine, as it is not always clear whether extraneous variables have an influence on the data being collected.

## INTERFERENCE

It is difficult to think of any ethical measurement of people's performance that does not affect or change the participants or their environment in some way, no matter how trivial. However, we would want to be sure that the reaction to the method and measurement process is not too great for the study purposes, and, critically, is known, acknowledged and reported. Rich participant observer accounts are often quite clear about the effect data gathering has had on the focus group, but the value of the data and interpretations frequently outweigh the scientific drawbacks. In an experimental setting, on the other hand, we would expect efforts to minimise any reactivity, to minimise bias and error and enhance reliability.

## SENSITIVITY

We would not want to use a wooden ruler to measure changes in stature due to exposure to vibration – it would just not be sensitive enough to the size of changes we expect. On the other hand, we would not want to go to the expense of recruiting 500 participants in an experiment when the effects we are interested in will be shown by almost all people and certainly will show up with 10 participants (e.g. some visual functions or performance). Methods and measures should be selected which are sensitive to the right degree to the sorts of effects we are examining, whether to confirm or reject our hypotheses. Related to this, we need to select measures which can provide the level of detail needed for the particular enquiry, but not too much greater than this (and with the range of new technologies and sensors that are now available, there is considerable temptation to 'over-measure' a situation).

## FEASIBILITY OF USE

Sometimes, it is said that methods should be simple to use – a little like product interfaces! This is true of methods which ergonomists have developed for use by others (engineers, managers, and health and safety specialists) in practical settings, but is not necessarily required for methods used by human factors and ergonomics specialists in laboratory or field. Some methods such as ethnographically informed observation, or specific techniques such as cognitive work analysis (Vicente, 1999 and see Chapter 3), require considerable training and experience in order to be able to apply them in a comprehensive form. However, we do want the methods used and the measures taken to be feasible for use in the particular circumstances we define, whether this is in relation to participants' understanding of questions, access points for observation to be set up, or enough data to populate a model meaningfully.

## ACCEPTABILITY AND ETHICS

Strongly related to the idea of face validity and also to feasibility, the methods and measures must be acceptable – to the client, the study population to whom they are applied and to our peers and colleagues in the wider scientific community. Such acceptance may be in terms of resources required,

interference with the activities of the population under study, or matters raised which the client would prefer not to be (e.g. worker opinions on matters outside the focus of the study). Very much related is the idea of ethical research and practice, where we should follow codes of conduct and guides to ethical practice such as produced by professional bodies (e.g. the Institute of Ergonomics and Human Factors or the UK Economic and Social Research Council). Key aspects of these are the way in which we treat our participants, how we represent ourselves and how we report on our findings (see Chapter 4 for further discussion of research ethics).

## RESOURCES

As with any endeavour, the resources required will be an important determinant of the methods and measures we use. Not only that but it will be the resources needed set against the importance of the study and the potential value of its outcomes that will be critical. Resources are not just to do with financial cost, although this is a central consideration; we must also assess the resources needed in terms of people (whether as investigators/researchers or as participants), equipment and also time – for the analysis as much as for the data collection, as is seen regularly in observation studies employing video analysis.

## CONCLUSIONS

E/HF has as its field the interactions of people with all other people and artefacts within an environmental context; its goal is the well-being of individuals, organisations and national economies. This requires appropriate theory and practice, models and methods. The thin line which we must tread when we become involved in discussion of methodology is summed up in two views: on the one hand, '[psychology] should not allow itself to be driven by obsession with method to the exclusion of the human problems that are its province' (Barber, 1988, p. 7, reporting Maslow, 1946); on the other hand 'anyone who wishes to reflect on how they practise their particular art or science, and anyone who wishes to teach others to practise, must draw on methodology' (Cross, 1984, p. vii). Certainly, ergonomists are driven by human problems, or driven to improve the quality of people's lives, but also we are always concerned to educate others (designers, engineers, politicians, accountants, managers, public, and media) in our approach and the necessity for it. If we truly believe that we are, above all, promoters of an approach and of a process, then it behoves us to pay great attention to the roots and current and future state of our methodology. The remainder of this book is but one step in doing this.

Earlier in this chapter, we explored frameworks for ergonomics and human factors, and in doing so suggested that the systems with which we should now be concerned cannot be modelled and understood by reference to a single person at the centre of a set of interactions with equipment and environment, but by representations of groups of people acting in social networks. In fact much work and life of the future, and many of the challenges facing ergonomics and human factors, will be found in complex interacting and distributed sociotechnical systems. The activities of interest are no longer just those of command and control, with individual operators working with single control display interfaces, but collaboration, co-ordination and integration, with constantly shifting and multiple teams of people, co-located and virtual, interacting with each other directly and via various forms of information and communication technologies (see Edwards and Wilson, 2004).

Because this focus for ergonomics is increasingly important in the future, then we will need to extend our range of methods to deal with it. We certainly will need to draw from fields such as ethnography (e.g. Hammersley and Atkinson, 1995; Engström and Middleton, 1996), social network analysis (Wasserman and Faust, 1994), and interaction design (Rogers et al., 2011), and utilise the best new developments in cognitive work analysis and design (e.g. Hollnagel, 2003). We continue to be in an exciting time for ergonomics methods in this area, and it will be a considerable challenge to find ways in which we can truly understand peoples' behaviour in such rich distributed

sociotechnical systems, and at the same time provide data and findings with enough rigour to be applicable in future design and implementation.

This book is about methods in ergonomics and human factors. In introducing the reader to how we study people's interactions with artefacts, environments and other people and how we gather information, form models, develop recommendations and designs and enhance understanding, we are also introducing them to ergonomics and human factors in general. Many other books, as noted throughout this chapter, do this as well, to different levels of detail, and in doing so introduce human factors methodology. The aim of *this* book is to discuss and consider the goals of different methods in different work contexts, to understand how we can best combine and apply the range of tools and techniques that we have available in a manner that ensures that we effectively support design of work systems and tools.

## REFERENCES

- Annett, J. (2002a). Subjective rating scales: Science or art? *Ergonomics*, **45**, 966–987.
- Annett, J. (2002b). A note on the validity and reliability of ergonomics methods. *Theoretical Issues in Ergonomics Science*, **3**, 228–232.
- Bainbridge, L. (1987). The ironies of automation. In: J. Rasmussen, K. Duncan and J. Leplat (eds.), *New Technology and Human Error*. London, U.K.: John Wiley.
- Barber, P.J. (1988). *Applied Cognitive Psychology*. London, U.K.: Methuen.
- Beatty, D. (1995). *The Naked Pilot*. Shrewsbury, U.K.: Airlife Pub. Co.
- Becker, B. and Gerhart, B. (1996). The impact of human resource management on organizational performance: Progress and prospects. *Academy of Management Journal*, **39**(4), 779–801.
- Bernard, H.R. (ed.). (1995). *Research Methods in Anthropology: Qualitative and Quantitative Approaches*, 2nd edn. Walnut Creek, CA: Altamira Press.
- Beynon-Davies, P. (1999). Human error and information systems failure: The case of the London ambulance service computer-aided dispatch system project. *Interacting with Computers*, **11**, 699–720.
- Bignell, V., Peters, G. and Pym, C. (1977). *Catastrophic Failures*. Milton Keynes, U.K.: Open University Press.
- Casey, S. (1993). *Set Phasers on Stun*. Santa Barbara, CA: Aegean Pub. Co.
- Chapanis, A. (1976). Keynote address: Ergonomics in a world of new values. *Ergonomics*, **10**(3), 253–268.
- Chapanis, A. (1979). Quo vadis, ergonomia. *Ergonomics*, **22**, 595–605.
- Chapanis, A. (1996). *Human Factors in Systems Engineering*. New York: John Wiley.
- Corlett, E.N. (1991). Some future directions for ergonomics. In: M. Kumashiro and E.D. Megaw (eds.), *Towards Human Work: Solutions to Problems in Occupational Health and Safety*. London, U.K.: Taylor & Francis Group.
- Cross, N. (1984). *Developments in Design Methodology*. Chichester, U.K.: John Wiley.
- Cushing, S. (1994). *Fatal Words*. Chicago, IL: University of Chicago Press.
- Dane, F.C. (1990). *Research Methods*. Belmont, CA: Wadsworth.
- Denzin, N. (1970). *Sociological Methods*. New York: McGraw-Hill.
- Denzin, N.K. and Lincoln, Y.S. (eds.). (1998). *Strategies of Qualitative Inquiry*. London, U.K.: Sage Publications.
- Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W.S., Wilson, J.R. and van der Doelen, B. (2012). A strategy for human factors/ergonomics: Developing the discipline and profession. *Ergonomics*, **55**(4), 377–395.
- Eason, K.D. (1997). Understanding the organizational ramifications of implementing information technology systems. In: M. Helander, T. Landauer and P. Prabhu (eds.), *Handbook of Human-Computer Interaction*, 2nd edn. Amsterdam, the Netherlands: Elsevier.
- Edholm, O.G. and Murrell, K.F.H. (1973). *The Ergonomics Research Society: A History 1949–1970*. Loughborough, U.K.: The Ergonomics Society.
- Edwards, A. and Wilson, J.R. (2004). *Guide to Implementation of Virtual Teamworking*. London, U.K.: Gower Press.
- el Kaliouby, R. and Robinson, R. (2005). Real-time inference of complex mental states from facial expressions and head gestures. In: *Real-Time Vision for Human-Computer Interaction*. Berlin: Springer-Verlag, pp. 181–200.
- Engström, Y. and Middleton, P. (eds.). (1996). *Cognition and Communication at Work*. Cambridge, U.K.: Cambridge University Press.

- Frankfort-Nachmias, C. and Nachmias, D. (1996). *Research Methods in the Social Sciences*. London, U.K.: Arnold.
- Green, B. and Jordan, P.W. (1999). The future of ergonomics. In: M.A. Hanson, E.J. Lovesey and S.A. Robertson (eds.), *Contemporary Ergonomics*, '99. London, U.K.: Taylor & Francis Group, pp. 110–114.
- Grey, S.M., Norris, B.J. and Wilson, J.R. (1987). *Ergonomics in the Electronic Retail Environment*. Slough, U.K.: ICL (UK) Ltd.
- Guardian. (2014). OMG! Number of UK text messages falls for first time. Published online 13 January 2014.
- Guion, L.A., Diehl, D.C. and McDonald, D. (2011). *Triangulation: Establishing the Validity of Qualitative Studies*. Gainesville, FL: Department of Family, Youth and Community Sciences, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, FCS6014, at <http://edis.ifas.ufl.edu/pdffiles/FY/FY39400.pdf>. Accessed 24th March 2014.
- Hammersley, M. and Atkinson, P. (1995). *Ethnography*, 2nd edn. London, U.K.: Routledge.
- Helander, M.G. (1999). Seven common reasons not to implement ergonomics. *International Journal of Industrial Ergonomics*, **25**, 97–101.
- Hignett, S. (2014). Commentary on the paper by Heimrich Kanis entitled: 'Reliability and validity of findings in ergonomics research'. *Theoretical Issues in Ergonomics Science*, **15**(1), 50–54.
- Hignett, S. and Wilson, J.R. (2004b). Horses for courses – but no favourites. A reply to three commentaries. *Theoretical Issues in Ergonomics Science*, **5**(6), 517–525.
- Hignett, S.M. and Wilson, J.R. (2004a). The role for qualitative methodology in ergonomics: A case study to explore theoretical issues. *Theoretical Issues in Ergonomics Science*, **5**(6), 473–493.
- Hollnagel, E. (ed.). (2003). *Handbook of Cognitive Task Design*. Boca Raton, FL: CRC Press.
- Hollnagel, E. and Woods, D.D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton, FL: CRC Press.
- Hollnagel, E. (2014). Safety-I and Safety-II The Past and Future of Safety Management. Farnham, UK: Ashgate.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Jian, J.-Y., Bisantz, A.M. and Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, **4**(1), 53–72.
- Johnson, G.I. and Wilson, J.R. (1988). Future directions and research issues in ergonomics and advanced manufacturing technology (AMT). *Applied Ergonomics*, **19**(1), 3–8.
- Kanis, H. (2014). Reliability and validity of findings in ergonomics research. *Theoretical Issues in Ergonomics Science*, **15**(1), 1–46.
- Kantowitz, B.H. (1992). Selecting measures for human factors research. *Human Factors*, **34**, 387–398.
- de Keyser, V. (1992). Why field studies? In: M. Helander and M. Nagamachi (eds.), *Design for Manufacturability: A Systems Approach to Concurrent Engineering and Ergonomics*. London, U.K.: Taylor & Francis Group, pp. 305–316.
- Kim, S.-W., Drury, C.G. and Lin, L. (2001). Ergonomics and quality in paced assembly lines. *Human Factors and Ergonomics in Manufacturing*, **11**(4), 377–382.
- Kragt, H. (1992). *Enhancing Industrial Performance*. London, U.K.: Taylor & Francis Group.
- Kuorinka, I. (ed.). (2000). *History of the International Ergonomics Association*. Boca Raton, FL: IEA Press.
- Mack, Z. and Sharples, S. (2009). The importance of usability in product choice: A mobile phone case study. *Ergonomics*, **52**(12), 1514–1528.
- Meister, D. (1985). *Behavioural Analysis and Measurement Methods*. Chichester, U.K.: John Wiley.
- Meister, D. (1995a). Divergent viewpoints: Essays on human factors questions. Personal publication.
- Meister, D. (1995b). Human factors – the early years. *Proceedings of the Human Factors Society 39th Annual Meeting*, San Diego, CA, pp. 478–480.
- Meister, D. (April 1998). Twenty-first-century challenges to ergonomics. *Ergonomics in Design*, **6**, 33–34.
- de Montmollin, M. (June 1992). The future of ergonomics: Hodge Podge or new foundation? *Le Travail Humain*, **55**(2), 171–181.
- Moray, N. (1994). “De Maximum non Curat Lex” or how context reduces science to art in the practice of human factors. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, Nashville, TN, pp. 526–530.
- Moray, N. (2000). Culture, politics and ergonomics. *Ergonomics*, **43**(7), 858–868.
- Muir, H.C., Bottomley, D.M. and Marrison, C. (1996). Effects of motivation and cabin configuration on emergency aircraft evacuation behavior and rates of egress. *The International Journal of Aviation Psychology*, **6**(1), 57–77.
- National Audit Office. (2011). The failure of the firecontrol project. HC 1272 Session 2010–2012. Department for Communities and Local Government. <http://www.nao.org.uk/wp-content/uploads/2011/07/10121272es.pdf>. Accessed 24th March, 2014.

- National Research Council. (1986). Research needs for human factors. Report of the Committee on Human Factors. Washington, DC: National Academy Press.
- Neale, V.L., Dingus, T.A., Klauer, S.G. and Sudweeks, J. (2005). An overview of the 100-car naturalistic study and findings. Paper 05-0400. National Highway Traffic Safety Administration. <http://www-nrd.nhtsa.dot.gov/pdf/ESV/esv19/05-0400-W.pdf>. Accessed 24th March, 2014.
- Norman, D. (1988). *The Psychology of Everyday Things*. New York: Basic Books.
- Oxford English Dictionary. (2014). [www.oed.com](http://www.oed.com). Accessed 24th May, 2014.
- Perrow, C.B. (1984). *Normal Accidents: Living with High Risk Technologies*. New York: Basic Books.
- Petroski, H. (1992). *To Engineer Is Human: The Role of Failure in Successful Design*. New York: Vintage Books.
- Pheasant, S. (1986). *Bodyspace*. London, U.K.: Taylor & Francis Group.
- Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L. and Morrisroe, G. (2005). The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload. *Applied Ergonomics*, **36**(6), 681–693.
- Poulton, E.C. (1975). Observer bias. *Applied Ergonomics*, **6**(1), 3–8.
- Proctor, R.W. and Zandt, T.V. (2008). *Human Factors in Simple and Complex Systems*, 2nd edn. Boca Raton, FL: Taylor & Francis Group.
- Reason, J. (1990). *Human Error*. Cambridge, U.K.: Cambridge University Press.
- Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Aldershot, U.K.: Ashgate.
- Robson, C. (2011). *Real World Research*. Oxford, U.K.: Blackwells.
- Rogers, Y., Sharp, H. and Preece, J. (2011). *Interaction Design: Beyond Human-Computer Interaction*. Chichester, U.K.: Wiley & Sons.
- Rouse, W., Kober, N. and Mavor, A. (eds.). (1997). *The Case for Human Factors in Industry and Government, Report of a Workshop*. Washington, DC: National Academy Press.
- Rutherford, A. and Wilson, J.R. (1992). Searching for the mental model in human-machine systems. In: Y. Rogers, A. Rutherford and P. Bibby (eds.), *Models in the Mind: Perspectives, Theory and Application*. London, U.K.: Academic Press, pp. 195–223.
- Scaife, M. and Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, **45**, 185–213.
- Shackel, B. (July 1996). Ergonomics: Scope, contribution and future possibilities. *The Psychologist*, **9**(7), 304–308.
- Sharples, S., Millen, L., Golightly, D. and Balfe, N. (2011). <http://www.swetswise.com/common/images/spacer.gif>The impact of automation in rail signalling operations. *Journal of Rail and Rapid Transit. Proceedings of IMechE, Part F*, **225**(2), 179–191.
- Singleton, W.T. (ed.). (1982). *The Body at Work: Biological Ergonomics*. Cambridge, U.K.: Cambridge University Press.
- Singleton, W.T. (July 1994). From research to practice. *Ergonomics in Design*, **2**(3), 30–34.
- Sparks, E.A., Wehbe-Janek, H., Johnson, R.L., Smythe, W.R. and Papaconstantinou, H.T. (November 2013). Surgical safety checklist compliance: A job done poorly! *Journal of the American College of Surgeons*, **217**(5), 867–873.
- Stanton, N.A. (2014). Commentary on the paper by Heimrich Kanis entitled ‘Reliability and validity of findings in ergonomics research’: Where is the methodology in ergonomics methods? *Theoretical Issues in Ergonomics Science*, **15**(1), 55–61.
- Stockbridge, H.C.W. (1989). *The Ergonomics Society: A History 1971–1989*. Loughborough, U.K.: The Ergonomics Society.
- Torgerson, W.S. (1958). *Theory and Methods of Scaling*. New York: Wiley.
- UK Government. (2012). Report of the service inquiry into the grounding of HMS Astute on 22 October 2010. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/271118/astute\\_grounding\\_si\\_report.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/271118/astute_grounding_si_report.pdf).
- UK National Audit Office. (27 October 1999). The passport delays of summer 1999. Report by the Comptroller and Auditor General. London, U.K.: The Stationery Office.
- UK Public Accounts Committee. (2000). Improving the delivery of government IT projects. Report of the House of Commons Public Accounts Committee. London, U.K.: The Stationery Office.
- Vicente, K.J. (1999). *Cognitive Work Analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Vink, P. (2014). Reaction on the paper ‘Reliability and validity of findings in ergonomics research’ of Kanis (2013). *Theoretical Issues in Ergonomics Science*, **15**(1), 47–49.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge University Press.
- Wastell, D.G. and Cooper, C.L. (1996). Stress and technological innovation: A comparative study of design practices and implementation strategies. *European Journal of Work and Organizational Psychology*, **5**, 377–397.



- Waterson, P.E. (2009). A systems ergonomics analysis of the Maudsone and Tunbridge Wells infection outbreaks. *Ergonomics*, **52**, 1196–1205.
- Waterson, P.E. (2011). World War II and other historical influences on the formation of the Ergonomics Research Society. *Ergonomics*, **54**(12), 1111–1129.
- Waterson, P.E. and Sell, R. (2006). Recurrent themes and developments in the history of the Ergonomics Society. *Ergonomics*, **49**(8), 743–799.
- Webb, E.J., Campbell, D.T., Schwartz, R.D. and Sechrest, L. (1972). *Unobtrusive Measures: Non-reactive Research in the Social Sciences*. Chicago, IL: Rand McNally.
- Welford, A.T. (1976). Ergonomics: Where have we been and where are we going: I. *Ergonomics*, **19**, 275–286.
- Wickens, C.D., Hollands, J., Banbury, S. and Parasuraman, R. (2013). *Engineering Psychology and Human Performance*, 4th edn. Boston, MA: Pearson.
- Wilson, J.R. (1994). Devolving ergonomics: The key to ergonomics management programmes. *Ergonomics*, **37**, 579–594.
- Wilson, J.R. (2000). Fundamentals of ergonomics. *Applied Ergonomics*, **31**, 557–567.
- Wilson, J.R. (2014). Fundamentals of systems ergonomics/human factors. *Applied Ergonomics*, **45**, 5–13.
- Wilson, J.R., Jackson, S. and Nichols, S. (2003). Cognitive work investigation and design in practice: The influence of social context and social work artefacts. In: E. Hollnagel (ed.), *Handbook of cognitive task design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wilson, J.R. and Nichols, S.C. (2002). Measurement in virtual environments: Another dimension to the objectivity/subjectivity debate. *Ergonomics*, **45**(14), 1031–1036.
- Wogalter, M.S., Hancock, P.A. and Dempsey, P.G. (1998). On the description and definition of human factors/ergonomics. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, pp. 671–674. Santa Monica, CA.
- Zhang, L., Helander, M.G. and Drury, C.G. (1996). Identifying factors of comfort and discomfort in sitting. *Human Factors*, **38**(3), 377–389.

# *Section I*

---

## *Fundamental Approaches and Tools*

The first section of this book describes a set of general approaches and methods that might be employed in any ergonomics/human factors (E/HF) enquiry or investigation, whether research driven or practice led. These methods help us to improve our fundamental understanding of people, whether in a laboratory or field context, and generally focus on capturing their activities, responses, reactions and thoughts. They might fit with an empirical approach, where we collect quantitative or qualitative data to record behaviours or opinions, or an analytical approach, where we create models or abstractions of people and what they do.

Many of the terms used within this field are heavily loaded, and there has been detailed and careful discussion about the titles of some of the chapters within this section. Distinctions such as direct/indirect or subjective/objective have been made in the past to classify methods, but these terms can be confusing or problematic in implying some element of ‘bias’ in subjectivity or interpretation of data. Broadly, the E/HF practitioner should normally be able to select the method appropriate for the question or situation, and be aware of the opportunities and limitations associated with that method.

The methods presented here include those which enable direct observation or capture of behaviours and interactions, and those which are indirect, and thus rely on interpretation from either the participant themselves (e.g. in terms of the way in which they respond to a question using a rating scale) or the experimenter (when classifying a behaviour whilst using a structured observation framework).

Although the direct observation group of methods is sometimes seen as ‘objective’ and somehow ‘true’, it is important to note that even data collected by direct observation will be re-analysed, summarised and abstracted by the observers, from their memory, notes, tape or video recordings, and then interpreted by the investigators, introducing some possible bias or at least subjectivity. This is even becoming apparent in very recent work which considers ‘big data’ analysis, where very large data sets, collected by, for example, commercial organisations, or technology-led companies such as Google, provide the apparent opportunity to no longer need to ‘sample’ and instead apply analysis approaches to the ‘complete population’. This needs to be done with caution – even in these data sets drawn from exceptionally wide sources, we cannot be confident that the complete population has been captured, and we should not forget the value of the statistical approaches that have underpinned our interpretation of experimental data for many years (as discussed in Chapter 2).

Similarly, a well-designed and implemented ‘subjective’ tool can elicit extremely valuable and reliable reports that simply could not be captured using ‘objective’ means.

What remains of course is the need for any E/HF practitioner to understand fully the nature of the method that they are applying, its limitations and the opportunities for the insight that it provides. The methods presented in this section enable readers to understand the underpinning considerations in selection of these broad methodological approaches, which can then be applied in a range of contexts, from laboratory to field.

Whilst much of the focus of this book is on methods that can be applied in the field context, it remains important for the E/HF practitioner to be aware of the fundamentals in experimental design and analysis. Controlled experiments allow us to isolate individual variables and, if extraneous variable are appropriately managed and robust measures applied, infer causative relationships between independent and dependent variables. Chapter 2, by Drury et al., presents a detailed description of the considerations involved in experimental design and analysis, with particular focus on the statistical consequences of different experimental design and analytical approaches. Whilst much of our work as E/HF practitioners is focussed in the field, or in tasks where it is simply too complex to control all of the influential variables, data to inform much of the underlying theory that informs our thinking, whether it is our response to an environmental variable, or the physical effort exerted in a task, is most effectively captured in a laboratory setting. Laboratory approaches allow us to test repeatability of results and understand the response of variables in isolation. So that such studies and experiments do not end up as costly, sterile exercises, great care must be taken in their design and conduct. The trade-off – being able to isolate the variables of greatest interest but by doing so perhaps losing the influence of setting and context – must be understood and assessed carefully before any choice to carry out experimental research.

Chapter 3 presents a consideration of the capture of complex, cognitive work. Whereas many of the later chapters in this volume use examples of physical and physiological responses, which can often be directly measured by instrumentation, this chapter focusses on methods used to capture and interpret tasks that are primarily cognitive. Such tasks are inherently harder to observe, and rely on a knowledgeable and experienced researcher capturing and analysing data that has been collected. Methods considered include structured interview approaches such as critical decision method, observational methods and generative methods, that are particularly useful when establishing user needs.

The general considerations that should be made when applying methods to capture or elicit responses from individuals more generally are presented in Chapter 4. In this chapter, a range of different methods and specific guidance to their design and approach are presented. This chapter covers a very wide range of general ergonomics techniques, from rating scale and questionnaire design, to developing interview protocols, to the ethical considerations of applying methods to capture and observe participant responses. A number of phenomena encountered whilst applying these methods, such as participant bias, are described, and approaches to minimise the effects of these biases on data collected are suggested. In addition, the psychophysics approach is described in this chapter. This is a method that underpins much of our understanding of how people perceive different sensations, and provides a specific approach to the quantification of responses. It is included in this book partly to ensure it does not become a ‘lost art’ but also to demonstrate the value and requirements of such a structured approach in eliciting specific responses from participants.

A specific chapter on qualitative methods was first introduced in the third edition of this book; the use of such methods now underpins much E/HF work both in research contexts and real world practice. Whilst in some situations, there is still a preference for, or perhaps confidence in, data that can be reported in a quantitative form, and the notion of statistical significance is still particularly important, the value of qualitative data, particularly when explaining *why* or *how*, is increasingly acknowledged. Hignett and McDermott, in Chapter 5, present the different philosophies associated with applying qualitative methods, and consider the different techniques that can be used to support such approaches.



A method that remains core to E/HF is task analysis (Chapter 6). General opinion within the discipline emphasises that task analysis is, as the name suggests, not just description but also analysis of tasks; within task analysis, data are collected, represented in an appropriate description form, and are analysed to assess task requirements for the person, their expected behaviour, any constraints and the task and environmental demands on them.

Task analysis is widely used, since it can be applied during the analysis of existing systems, the design of new ones and the evaluation carried out subsequently. Information gained is useful both in development and as criteria against which to assess what is developed. What is represented and analysed is what must be done in order to fulfil certain goals, within constraints from the task environment and from individual or general human limitations. Task analysis and its techniques were originally distinguished from method study both by underlying purpose but also by their concentration upon operator decisions as much as upon actions. This chapter also touches upon user modelling approaches, which use outputs of task analysis as the basis for predictive models of user performance to inform and test prospective designs.

One particular goal of task analysis and analysis of cognitive work is to understand the nature, manifestation and elicitation of knowledge, taken up by Shadbolt and Smart in Chapter 7. Human factors have a role to play in addressing the bottleneck in knowledge-based systems – acquiring the knowledge necessary to build the system. Generally, the major source of such knowledge will be the experts themselves; knowledge elicitation is the name for gathering the relevant information about people's knowledge. This chapter also describes semantic web approaches to representation of knowledge elicited from individuals in machine-readable form, moving beyond dedicated 'expert systems' to a generic approach that allows knowledge to be captured and stored in a structured and searchable manner. In some ways, parallels can be drawn between this and chapters on involving people in design research and participatory ergonomics, later in this volume, in that the techniques described there are to elicit information, opinion and insight from users or user 'representatives', to be applied in design.

Eliciting expertise uses a number of methods from other domains but has also seen development of a number of methods explicitly for the purpose of understanding the knowledge and skills that people hold, which is why this chapter appears in this general section. Moreover, the focus is now on the support for knowledge management, knowledge-based systems of all kinds and for design of jobs and organisations that make the most of the skills and knowledge of the staff.

Finally, in Chapter 8, the approach of user modelling in simulation and digital human modelling forms is discussed. Many tools now exist to represent users and their potential work contexts in a three-dimensional form – if designed and applied appropriately, they provide a rich resource to analyse the impact of a design intervention or modification earlier in the design phase than would previously have been possible. However, as with other methods, it is vital that a practitioner is aware of the limitations of these approaches; in the case of digital human modelling for example, a model is only going to be as good as the source from which anthropometric data is drawn and the biomechanics built within the computer representation. If the underpinning representation is 'fit for purpose' (i.e. it has the accuracy and granularity required for the topic under question), then this can be a very powerful and valuable predictive technique, not only for analysing the cognitive, physical and environmental 'fit' of a future design, but also for communicating the output of an E/HF analysis to key stakeholders and decision makers.

A message throughout this book is that methods and techniques are only tools, to be selected and applied with a clear understanding of the design or evaluation objectives. However, although skilful use of methods requires practice, we should beware of sticking rigidly to standard methods. Imagination, as well as scientific rigour, as stated in the third edition of this text, remains a key ingredient of successful ergonomics.



---

# 2 Experimental Design and Analysis

*Colin Drury, Victor Paquet and Harrison Kelly*

## CONTENTS

|                                                            |    |
|------------------------------------------------------------|----|
| Introduction.....                                          | 37 |
| What Is an Experiment?.....                                | 38 |
| Factors and Levels.....                                    | 39 |
| Effects .....                                              | 42 |
| Hypotheses.....                                            | 43 |
| Experimental Design Alternatives .....                     | 45 |
| The Basis of Factorial Experimentation and Analysis .....  | 45 |
| Main Effects .....                                         | 46 |
| Interactions.....                                          | 47 |
| Design of Multifactorial Experiments.....                  | 48 |
| Within-Participant versus Between-Participant Designs..... | 49 |
| Sample Size, Effect Size and Power .....                   | 51 |
| Fractional Factorial Designs .....                         | 52 |
| Sequential Experimentation .....                           | 54 |
| Analysis Alternatives .....                                | 55 |
| Dealing with Data .....                                    | 55 |
| Dealing with Assumptions .....                             | 56 |
| Dealing with Analysis Packages .....                       | 57 |
| References.....                                            | 58 |

## INTRODUCTION

This chapter provides the ergonomics/human factors (E/HF) practitioner and researcher with practical advice on how to design studies and analyse the resulting data to achieve effectiveness and efficiency. In the 1930s, there was a major revolution in how experiments are designed. We moved from the traditional physical sciences model of ‘vary one factor at a time; keep all other factors fixed’ to a philosophy that emphasises varying multiple factors in the same experiment. What made this revolution possible was the development of sophisticated statistical techniques (e.g. analysis of variance [ANOVA]) that allowed for the parsing of the effects of each factor, and their combinations, and the testing of each effect against the normal variation experienced in any experiment. This is the model for experimental design and analysis we still use in E/HF.

Within this chapter, we use a number of terms as if they were already understood by researchers and practitioners. Examples are ‘experiment’, ‘factor’, ‘effect’ and ‘statistical techniques’. We can do this because most E/HF professionals have had courses with titles such as ‘design of experiments’ (DoE) at some point in their careers. This chapter will cover some familiar ground, such as multifactorial experiment designs and analysis of variance, but will hopefully extend the topic in a number of ways. First, we will provide some definitions that broaden the topic to other E/HF studies. Next, we will delineate statistical techniques for data that does not fit the usual normality

assumptions, and finally, we will add material on power calculations, effect magnitudes and meta-analysis that will at least provide sources for further reading.

## WHAT IS AN EXPERIMENT?

An experiment is ‘a scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact’ according to the Oxford Dictionary ([www.oxforddictionaries.com](http://www.oxforddictionaries.com), 2014). Experiments differ from all other techniques in that they directly change a system, normally in the form of a manipulation of the independent variables. These independent variables are derived from research hypotheses and enable the measurement of the degree to which each independent variable impacts and affects the dependent variable of interest. The research hypotheses are usually stated in the form of a null hypothesis, which states that the independent variable has no effect, and an alternative hypothesis, which states that the independent variable has an effect. The convention is to use statistical results to either ‘reject’ or ‘fail to reject’ the null hypothesis with a degree of certainty (sometimes also reported as supporting the alternative hypothesis). Hypotheses are discussed in more depth later in this chapter.

Alternative explanations of the results in terms of coincidence are most unlikely, because the experimenter determined when and how to change the system. Thus, experiments are able to detect and infer causality with some degree of confidence. A typical designed experiment is that of Drury et al. (2008) who studied the effects of workplace posture on performance and comfort in a security screening task. They used three postures at the airport workplace for screeners viewing x-ray images of baggage potentially containing threat objects. Screeners could be either standing, sitting on a high chair or sitting on a normal-height chair. The design of the experiment used three factors: posture, run order (to measure any learning/fatigue) and participant (12 experienced screeners). (The findings were no effects of posture on performance, but large effects of posture on discomfort and of run order on performance.) The results suggest that ‘posture’ can affect discomfort and ‘order’ can affect learning, but no conclusions can be reached about the effects of ‘posture’ on performance or ‘order’ on discomfort.

Experiments are not the only form of E/HF enquiry, as other chapters in this book amply demonstrate. However, they do have the most obvious impact on determining causality. Also, the same statistical design techniques used in experiments can also be used in other forms of study such as observation studies (where, as discussed in Chapter 1, we can sometimes impose an ‘experimental paradigm’). We can choose which factors and which levels of these factors to observe, and apply similar statistical techniques to the data to understand the separate effects. As an example, Chang and Drury (2007) studied human interaction with doors by observing 1600 people as they used different doors. The doors (both push and pull types) were chosen to give four levels of physical difficulty, characterised by the restoring torque in N m needed to move the door. The people observed were characterised by their strength, determined by their gender and stature (observed against a set of marks on the door). The measurement was the use of body weight, rather than just one hand, to help open the door. Even though this was an observational study rather than an experiment, a statistical multifactorial design was used, one recognisable by anybody taking a DoE course. (Chang and Drury found, as expected, that task demands [door difficulty] and human capability [stature used as a proxy for upper body strength] jointly determined the likelihood of using body weight to open the door.)

As noted in Chapter 1, an experiment is NOT automatically a laboratory study. While laboratory studies are often experiments, so are field studies. For some classes of enquiry, such as examining the details of visual search or biomechanics, a laboratory may be the most convenient place to control the factors of interest and measure their effects precisely. But a valid experiment can take place in a field setting, for example measuring the effects of changes in workplace design or aircraft cockpit advances. In a typical field study, Latorella and Chamberlain (2001) asked experienced pilots to fly a twin-turbo-prop aircraft towards convective weather fronts to measure the effectiveness of three different weather displays on the pilots’ decision points for avoiding the weather front. Note that the

experimenters made deliberate changes to the system (three displays, nine quantised distances from the weather front) and were in a position to measure both the pilots' decisions and their responses to weather questionnaires. This was not a simplified laboratory simulation, although much prior research derived from laboratory simulations was used in the detailed design of the experiment.

Any experiment changes the system under study. Thus, it is reactive on the system. Participants know the experiment is taking place and most details of how it will affect them. In any academic or research institution or government organisation, an Institutional Review Board (IRB) or ethics committee will have vetted the study before it is allowed to proceed. This helps ensure that the experiment is ethically designed and executed, protecting all participants to the extent possible, but it also emphasises that an experiment does impact the system and its participants. An IRB review might not always be used in an enterprise setting (business, industry) where the enterprise employs the experimenters and participants, although it should always be the norm.

Direct intervention can be both costly and potentially dangerous. Given that, what does this reactivity buy for the E/HF professional? In addition to the major advantage of determining causality, the three other advantages of a more highly reactive design are as follows:

1. The ability to be in the right place at the right time to observe. This is particularly important in human factors studies where the system behaviour observed is rare and unexpected (e.g. accidents or breakdowns).
2. The ability to use more obviously invasive, but information-rich, measurement techniques. For example, in inspection research, the response to each individual item inspected can be observed in considerable detail in an experiment (e.g. Drury and Sinclair, 1983), whereas in the real situation, only a simple accept/reject response is often given.
3. The ability to control or manipulate other variables not of primary interest that would otherwise alter the results of the investigation. Variables that are unaccounted for and contribute to variability in the measured effects of a study can mask the effects of the independent variable. In cases where such unaccounted variables are also correlated with the independent variable, the results of a study become 'confounded' by the unaccounted-for or spurious variable and it is impossible to determine the relative contributions of the independent variable and the confounding variable. Experiments, when designed properly, account for these variables so as to limit the spurious variability of the effects and minimise confounding. For example, in the security screening study, the order of presentation of conditions to screeners would be likely to have an effect, so this variable was controlled by explicitly including it as a factor in the study.

If the E/HF professional gains in experimental control and measurement detail by using highly reactive designs, what is lost? The major loss is in face validity. If we observe a system in its natural state, those associated with the system and possibly those who commissioned the study can be convinced that the study is realistic. An experiment, particularly one performed in a laboratory with artificial stimuli and non-representative participants, requires much more persuasion on the part of the E/HF professional to gain acceptance. The lead author was once involved in two studies of fork-lift truck control. One (Drury and Dawson, 1974) involved real drivers using real fork-lift trucks in a real warehouse to study lateral control behaviour. The other (Drury et al., 1974) involved real drivers controlling a toy train in a laboratory to study longitudinal control behaviour similar to Fitts' law tasks. It is obviously much easier to quote the former study to convince warehouse managers of its design implications.

## FACTORS AND LEVELS

What scientists refer to as independent variables are characterised in DoE as factors. They are the things the experimenter varies in order to measure their effects. In the doors study example, the factors were 'physical difficulty', 'gender', 'push vs. pull' and 'stature'. For the weather display study,

the factors were 'display', 'proximity to weather front' and 'participant'. In the security screening study, the factors were 'posture', 'run order' and 'participant'. These are examples of factors designed into the experiment with specific values, called levels. For example, in the doors study, the factor 'gender' was at two levels: male and female. The factor 'physical difficulty' had four levels of door restoring torque: 30, 46, 55 and 72 N m. In the weather display study the factor 'display' was at three levels: aural only, out-the-window view plus aural and a graphic weather display plus aural. In that study also, there were six participants and the factor 'proximity to weather front' was at six levels: every 20 nautical miles from 120 down to 20. In the security screening study, 'posture' was at three levels: standing, high chair and normal chair, and so on. The combinations of levels studied in an experiment are called experimental conditions.

Statistical DoE emphasises factors and levels, almost to the exclusion of other ways of treating independent variables. In any study, there are a potentially infinite number of variables that could possibly affect the outcome. To reduce the normal variation between data points, the E/HF professional must deal with all of these variables, even though that sounds impossible in theory. Some independent variables become part of the experiment by including them as factors at a number of levels ( $>1$ ), as in the earlier examples. The problem with allowing each variable to take on numerous levels is that in most experimental design solutions, the total number of trials is determined by the *product* of the number of levels of each factor. At the other extreme are what a physical scientist would call 'nuisance variables', which could affect the outcome if not controlled in some way.

The most useful ways of controlling these extraneous variables are to fix them at a single level or to use random assignment so that they do not *systematically* bias the outcome and conclusions. Fixing a variable at a single level is the most obvious way to control it. In the doors example, the whole study took place on a campus, eliminating some variability in age. For the weather display study, all participants had to be instrument-rated pilots with minimum experience defined by specific numbers of flights and flight hours to eliminate much variability due to skill and experience. Trained security screeners were used in the security study for the same reason. There is a price to be paid for fixing a variable at a single level in that the results may strictly apply only to that level of that variable. Thus, the doors study outcomes would not necessarily apply to, for example, people in senior citizen housing, in an elementary school or in a private home. In the weather display study, we could not generalise to, for example, novice pilots or military pilots.

Using randomisation to prevent, for example, all of the older participants being tested on a Monday is a powerful tool that prevents systematic bias by ensuring that any uncontrolled variability contributes only to the 'normal variation'. In statistical terms, this prevents bias to the mean, possibly at the cost of increasing the residual variance. Randomisation is a safe way to control unwanted variation, but the cost of the study might increase, because larger sample sizes are needed to reach the same level of certainty in the conclusions. The security screening study used three sets of baggage images to avoid screeners recognising particular bags: these three sets were presented in a different random order to each participant. A potential alternative to randomisation for reducing order effects (learning and/or fatigue) is to counterbalance the order in which factor levels are presented to each participant. It works best when any change over time or order is linear, which is rarely the case. Other order-balancing designs, such as Latin Squares, are presented later under DoE.

There is another way to treat an independent variable, part way between treating it as a factor at several levels and fixing or randomising it. If we know of an independent variable that is likely to affect the outcome of the experiment, we can measure that variable and treat it formally in the design as a covariate rather than a factor. As an example, in studies of inspection tasks such as that studied by Drury and Sinclair (1983), much evidence has been found that the cognitive ability of inspectors helps determine their effectiveness (Drury et al., 2009). One good measure is that of field independence, measured by an embedded figures test, first developed in the 1950s (Witkin, 1950; Jackson, 1956). Using the individual scores of field independence as a covariate to account



for the effects of different cognitive abilities when different groups of people are tested for different conditions removes some of the random error variance, thus increasing the power\* of the study.

The E/HF professional designing a study must trade off the increased power coming from the reduced variance due to fixing a variable at a single level against the decreased applicability of the results beyond that single level. There is no rule for this; it depends upon the aims of the study. For example we typically sample participants from some population of interest (e.g. supermarket workers, retired women or current air traffic controllers). The list of all such potential participants is called the sampling frame. There are correct statistical procedures both for sampling safely and for treating the results so that we can generalise to the whole population of interest. The basic choice is between systematic sampling and random sampling. We start with the sampling frame and use either a systematic method (e.g. every fifth person) or a random method (e.g. choosing via random number table or programs) to determine the members of our sample (e.g. Section 3.3.3.2 of the *NIST Statistics Handbook* <http://www.itl.nist.gov/div898/handbook/>). In general, randomisation is statistically safer, although stratified sampling by randomly choosing from different defined strata of the sampling frame can be useful where more specific results are required. Occasionally, we are only interested in the specific participants who have taken part, particularly where there is a very small population from which to choose (e.g. active cosmonauts). This was the case in the fork-lift truck control study referenced earlier (Drury and Dawson, 1974) where only four drivers had been trained on all of the trucks tested and no more training was planned. In this case, the four participants were technically a fixed factor (at four levels of course) rather than the usual random factor. This changed the statistical treatment later in the analysis.

We have laid out some of the statistical and design issues in choosing factors and levels. However, the main considerations are more practical (i.e. design or resource driven) than statistical. DoE texts (e.g. Winer, 2012) or online expositions (e.g. *NIST's Statistics Handbook* <http://www.itl.nist.gov/div898/handbook/>) give general guidance on choosing factors, but cannot be specific when aimed at a diverse audience. Much traditional teaching of DoE uses data-driven insight, as if the process were truly a black box that needed to be understood from existing data or preliminary studies. Thus, exploratory data analysis is seen as a first step in process understanding. In E/HF we can provide more specific guidance. The key is E/HF insight, that is what is important in a situation or process. There is no substitute for E/HF knowledge, whether from textbooks, journal papers or applying E/HF insight into a specific process through direct observation. E/HF has over half a century of accumulated knowledge, on top of hundreds of years of insight from component disciplines such as psychology, physiology or occupational epidemiology. In DoE, we ignore this knowledge at our peril.

E/HF insight means understanding which variables affect individuals' and systems' behaviour, health and performance, and having some feel for the relative magnitude of the effects of these factors. Some factors are obvious, even to lay persons: there will be differences between individual participants in the experiment, there will be learning effects when a task is relatively new to the participant, larger participants will often be stronger, and so on. Some are more obvious to an E/HF practitioner: Task overload will often lead to degraded performance, non-neutral postures held for long periods will result in discomfort, performance may be worse on night shifts or with insufficient sleep.

Within this, some variables are almost mandatory and unavoidable. For example individual differences between participants or temporal changes within the study (learning, fatigue) could affect behaviour, health or performance. These have been discussed earlier, but we return to them later under DoE. Unless we deal actively with such variables, our conclusions will potentially be flawed. We cannot just ignore such factors and hope they will not affect our results. Study sponsors and journal editors will both catch such design problems, at the point where it is too late to deal with them. Other factors of importance are dependent on the particular study. Physical ergonomics studies would

---

\* Power denotes the extent to which an experimental design is able to detect an effect that does indeed exist. A study with a power of 0.8 has an 80% chance of detecting the effect, if there is actually such an effect.

suggest that age, body size, strength and endurance are individual factors that could be important when assessing health (or physiological effects) and physical performance, and probably need to be dealt with by one of the techniques discussed earlier for reducing unwanted effects of individual differences. But other factors such as temperature, humidity and altitude would potentially be important too, as would workplace layout and task pace. We would expect some, but not huge, effects of cognitive variables on physiological effects and physical task performance, and no effects of such outlandish factors such as eye colour or moon phase. Similarly, cognitive ergonomics would suggest that age, cognitive style and task training would be very important to assess behaviour, health and performance during cognitive tasks, as would display layout strategies and task pacing. Physical factors would have less importance, but certainly measurable effects (see review in Drury et al., 2008).

In teaching experimental design, we have often found that the choice of sensible factors and number of levels is the most difficult aspect to communicate. Perhaps this is because traditional experimental design textbooks tend to ignore it, leaving more detailed knowledge to domain expertise. But the statistical and domain-specific aspects of DoE are rarely brought together in domain teaching (e.g. in E/HF subjects).

## EFFECTS

The basic classification of E/HF-relevant effects comes from the definition of Ergonomics provided by the *International Ergonomics Association* (IEA) as:

Ergonomics (or human factors) is the scientific discipline concerned with the understanding of the interactions among humans and other elements of a system, and the profession that applies theoretical principles, data and methods to design in order to optimize human well-being and overall system performance... (<http://www.iea.cc/>)

Clearly what E/HF professionals are ‘about’ are the twin groups of measures: system performance and human well-being. Just as in a road test of a car, there is no use knowing its speed and handling characteristics without also understanding its fuel consumption and reliability; in E/HF, we typically need to measure both the system performance and the cost to the human of achieving that performance.

‘Performance’ measures can involve the overall system, but are also possible for sub-systems even down to the individual human or group of humans. They can be simply classified into measures of speed and accuracy, or ‘time and errors’, although it is possible to argue that sub-standard speed performance is in itself an error so that we really only need to measure errors (e.g. Drury, 1994a). Speed measures are any measures with time in the numerator or denominator. In the numerator, they can be performance times, cycle times, reaction times at the individual, sub-system or system level, or even broader measures such as system down-time or systems availability. In the denominator, they are speed measures rather than time measures (e.g. output per shift, bits-per-second, rate of progress or miles per hour, as in driving). Accuracy is the positive aspect of performance (e.g. number of hits on target, quality level of a production process, percentage of driving task spent within the desired roadway). The negative aspect is errors (e.g. number of Methicillin-Resistant *Staphylococcus Aureus* [MRSA] infections per month at a hospital, fraction defective in a production process, percentage of missed threats and false alarms in a security process). Errors can either have a time-rate aspect (infections per month) or an event-rate aspect (fraction defective). If the time or event horizon is pre-specified and constant, perhaps the duration of the study, then the raw numbers of errors can be counted. As with the complementarity of performance and well-being measures, so speed and accuracy often need to be measured together. How can we praise the speed of a Formula 1 driver who has frequent crashes? How can we praise the output quality of a process operator who misses all deadlines?

Speed and accuracy do trade-off (although they may appear not to in current quality literature: Drury, 1997), often enough that many have studied the ‘speed-accuracy trade-off’ or SATO (e.g. Drury, 1994b). This is also known by the more accurate descriptor speed-accuracy operating



characteristic [SAOC] (e.g. Pew, 1969), because the plot of accuracy versus speed is in the form of a statistical operating characteristic (OC) curve. Typically, more accuracy demands less speed (e.g. in Fitts' Law; Hoffmann, 1992), or in visual search tasks (Drury and Forsman, 1996). Also different errors can trade off against each other, the most obvious example being correct detections (hits) against false alarms in a signal detection task (e.g. Fisher et al., 2012). Thus, choosing measures of performance needs good E/HF models of how the suite of performance measures might co-vary. It is assumed that there will be a suite of measures to capture both system and human performance (see Chapter 1 for the range of measures that might be selected).

Well-being measures can range from the 'soft', such as discomfort ratings, to the dramatic, such as fatalities. They tell of the positive and negative effects of the system, and the human role in the system, effects on the humans within the system, or even those whom the system impacts in use. Typical measures cover workload (e.g. NASA Task Load Index, TLX, scale), internal state of the operator (e.g. discomfort, fatigue), physiological/biomechanical stress, freedom from long-term diseases, health status, negative incidents (e.g. near misses), actual accidents and injury/equipment damage/fatalities. Most of these are covered elsewhere in this book (see Chapters 15, 18, 20 and 31 for examples).

The measures of effects must themselves measure up to scientific adequacy. No measure is useful unless it can be measured with sufficient reliability and actually represents a parameter of interest to those commissioning or reading the study (refer to Chapter 1 for a more in-depth discussion of reliability and other considerations when selecting methods). As with much of what is presented in this chapter, initial resolution of these issues with those who commission the study can save much grief when the study findings are finally presented.

## HYPOTHESES

In designing an experiment, or other E/HF study, a critical challenge is turning a research hypothesis into one or several statistical hypotheses to guide future actions. With the revolution in statistical DoE and statistical testing came a profound change in how we interpret results. The philosophy is that we pre-specify which outcomes of the study will lead to which conclusions. Thus, the E/HF professional could in theory give the designed experiment to a competent subordinate and be assured that whatever the data outcome, the conclusions will be exactly as planned. (This rather rigid statistical approach of course contrasts with more grounded theory or emergent approaches as are described in Chapters 3 and 5.) Such a rigorous scientific methodology is often followed until the last step of conclusions, where experimenters have been known to hedge their bets somewhat when faced with conclusions they do not like. For this reason, we spell out the transformation of a research hypothesis into a testable statistical hypothesis and provide an example of following through to conclusions.

The example is the study of security screening outlined earlier (Drury et al., 2008). This did not start out as a study of security screening, but as an examination of 'the inter-relationship between physical ergonomics and cognitive performance'. The experimenters were in fact a graduate class performing a rigorous E/HF practicum (Drury et al., 2007). The aim was to review the extensive literature on the interactions of physical and cognitive work, and test key aspects of this experimentally. (The idea of using security screeners did not arise until later, when it was found to be a convenient domain for testing the research hypotheses.)

The process involves several steps, adapted from the text by Siegel and Castellan (1988), which are illustrated here for the security screening study, concentrating on one of the hypotheses tested.

1. *State the research hypothesis:* Here the research hypothesis was that there is a measurable effect of the posture enforced by the physical workplace on performance in a cognitive task. That is not particularly new, as some authors have found such an effect in the past, but others have not (e.g. Mozrall and Drury, 1996).
2. *State the null and alternative hypotheses:* For a statistical test, we must turn the research hypothesis into a null hypothesis that states 'nothing was found'. The negation of this null

hypothesis is the alternative or experimental hypothesis. The data will eventually show which hypothesis we can conclude is true. The use of a null hypothesis is important as we can find the sampling distribution of any test statistic fairly easily when there is no effect. Here the null hypothesis was that there is no difference in any performance measures (hits, false alarms, time per bag screened) between the three workplace postures tested, that is 'there is no effect of the workplace postures on the performance measures'.

3. *Choose the statistical test:* In a multifactorial experiment such as the screener study (where the factors were posture and participant), the most logical test is the F-test for many cases when the effects are measured on interval or ratio scales. A test statistic (such as t or F) is a dimensionless quantity that is typically calculated as the ratio of the size of an effect to the appropriate variability. The test statistic increases in absolute magnitude as the size of the effect increases and as the variability of the effect within a condition decreases. Size of an effect is represented by the difference between two means (or the variance between several means for an F test) and is thus a fact, although subject to sampling error. The appropriate variability for a test statistic is the standard error of the difference between two means, which in simple cases is

$$SE = \frac{\text{Standard deviation}}{\sqrt{(\text{Number of data points})}} = \frac{SD}{\sqrt{N}}$$

where SD is the standard deviation of a set of N data points. Thus the test statistic is

$$\text{Test statistic} = \frac{M1 - M2}{SE} = \frac{(M1 - M2)\sqrt{N}}{SD}$$

where M1 and M2 are the means of the conditions being compared. For any statistical hypothesis, there are typically alternative tests with different assumptions. Non-parametric alternatives, which make fewer assumptions about the distribution, type and source of the measurement values, are possible. However, these tend to only be suitable for simple designs involving only one or two independent variables (e.g. Siegel and Castellan, 1988).

4. *Find the sampling distribution of the test statistic under the null hypothesis:* The F statistic's distribution is well known, tabulated in almost any statistics or DoE text and an integral part of most statistical software available for DoE and analysis. It does, however, rely on the assumption of normally distributed data, homogeneity of variance, independence of sampling and use of interval or ratio data. The assumptions and how violations of the assumptions impact results are also described in most DoE texts.
5. *Select the level of significance:* The level of significance or 'p value' is our threshold for concluding that the alternative hypothesis is true. This defines the likelihood of concluding that an effect exists when it truly does not exist, an erroneous conclusion. In statistical texts, this called a Type I error, and it is important in experimental research to ensure that this likelihood is small. In research, we typically choose a level which would rarely be found by chance (e.g. 1 in 10, 1 in 20, 1 in 100 or 1 in 1000). In this way, we limit our false alarm rate *when many tests are performed*. We do not eliminate false alarms, just make them rarer. If a researcher publishes 100 studies in a career, each with a single test at 1 in 20 ( $p = 0.05$ ), then about 5 false alarms would be expected. A couple of notes are in order here. First, there is no generally accepted p value – it depends upon the circumstances. A p value of 0.01 may be needed if the consequences of false alarm are very high. A p value as low as 0.1 may be acceptable when there is a limited population to collect data from, but a decision is still required. Second, in theory, the person who commissions the study should

choose the significance level. This is fine when the commissioner is a government, scientific or medical agency. However, most managers, most lawyers and many public servants are quite unused to the idea of probability and balk at choosing a level because of their lack of statistical knowledge. Here, E/HF researchers must help them reach an informed opinion, just as they must help a participant understand the risks involved in study participation. In industry, which should be quite used to probabilities after years of Six-Sigma programs, E/HF professionals still get asked ‘what is the right answer?’ when they attempt to involve managers in the choice of level of significance. Taking some responsibility to help our colleagues from other fields is in fact helping to fit the task to the decision maker. In the screener study, we chose a level of significance of  $p = 0.05$ , a very traditional value in science, and the normal minimum level of significance adopted in E/HF journal publications, because rejecting the null hypothesis with only a 5% chance of a ‘false alarm’ would convince others in our profession that we really had found an effect.

6. *Determine the region of rejection:* If we know the sampling distribution of the test statistic and the probability we will accept for a false alarm, then we can split those values of the test statistic into a set for which we will accept the null hypothesis and a set for which we will reject the null hypothesis. Thus, we have mapped all possible outcomes (values of  $F$  in our example) onto the conclusions we will draw before running the experiment. From here on, the process is purely mechanical. The region of rejection of the  $F$  test in the screener study was those values of  $F$  beyond which lay only the upper 5% of the probability distribution. The exact  $F$  value depended upon the number of degrees of freedom\* in the numerator (2 in this example, because there were three levels of the posture factor) and the denominator, which could vary, because there was more than one effects variable, and each had a different number of measurements.
7. *Run the study and determine the value of the test statistic:* This is the most time-consuming step, but collecting the data and calculating the  $F$  value are issues covered later. Setting up the screener experiment and collecting the data took several weeks of effort by all six authors of that study.
8. *Determine whether or not the data support the alternative hypothesis based on the criteria established in Step 6:* In the screener example, the null hypothesis was accepted for all performance measures.

What we have done is to force ourselves to pre-select our ultimate actions based on the statistical outcome. As with our discussion of factors, levels and the qualities of measures chosen, this discipline forces us as experimenters to think about the process of experimentation through to the final conclusions *before* we begin even recruiting participants or building equipment. In this way, there should be fewer studies that fail to meet the expectations of either client or experimenter.

## EXPERIMENTAL DESIGN ALTERNATIVES

### THE BASIS OF FACTORIAL EXPERIMENTATION AND ANALYSIS

Any standard DoE text (e.g. Winer, 2012) will provide literally hundreds of potential designs for statistical experiments, with detailed instructions on the choices available and the correct analysis techniques. Clearly, such a treatment is inappropriate here, so we shall concentrate on issues of most interest to the E/HF practitioner. Most texts start with simple comparisons of two levels of a single factor using a  $t$ -test or non-parametric equivalent. Because we are using people as participants in our experiments, and differences between participants are non-trivial, we are rarely able to use such basic designs. Unless we wish to run the study on a single participant, we always have one factor

---

\* Degrees of freedom are related to the number of values that can vary in a calculation.

of participant with differences between participants as a factor in our analysis. Even a comparison of two levels of a factor of interest will result in a multifactorial experiment, so that is where we must start.

Underlying all multifactorial designs is the analysis concept of ANOVA, again covered in depth in DoE texts. This in turn rests on the mathematical property of a variance: the variance of the sum of, or difference between, two means is equal to the sum of the variances of the separate means:

$$\text{Var}(x_a + x_b) = \text{Var}(x_a) + \text{Var}(x_b) \quad \text{and} \quad \text{Var}(x_a - x_b) = \text{Var}(x_a) + \text{Var}(x_b)$$

This simple assertion (again, proved in most statistical tests) extends to a linear combination of variables so that the variance of their sum is the sum of their variances. Thus, if we run a two-factor model with Factor A at a number of levels denoted by [i] and Factor B at levels [j], we can say that any single data point at level i of Factor A and level j of Factor B is composed of:

The overall mean of the whole data set

- + the difference between the overall mean and the true effect of Factor A at level i
- + the difference between the overall mean and the true effect of Factor B at level j
- + the random error of the combination of Factor A at level i and Factor B at level j

$$X_{ij} = \mu + A_i + B_j + \varepsilon_{ij}$$

This is the structural model of the experiment. Taking variances, which are additive, and noting that the variance of a constant such as  $\mu$  is zero, we have

$$\text{Var}(X_{ij}) = \text{Var}(A_i) + \text{Var}(B_j) + \text{Var}(\varepsilon_{ij}),$$

provided that the error variance is the same for all [i,j] combinations of Factor A and Factor B. In this way, we can take the overall variance of all of the data ( $\text{Var}(X_{ij})$ ) and split it into components due to Factor A ( $\text{Var}(A_i)$ ), Factor B ( $\text{Var}(B_j)$ ), and the error variance ( $\text{Var}(\varepsilon_{ij})$ ). We can thus see how much variability is uniquely associated with each factor, called here a ‘source of variance’. We can also compare this variance with the error variance to form a test statistic, as in the previous section, because a test statistic is an effect size divided by its appropriate variance. We have just partitioned variance into its components, so it is unsurprising that the technique is called ANOVA. If we can do it for two factors, then we can do it for any number of factors, provided that their effects are additive. Note that this paragraph has two uses of ‘provided that’, implying that to use ANOVA, we must make assumptions about homogeneity of variance and additivity of factor effects. We shall return later to alternative ways to analyse the data if these assumptions are not met.

Multifactorial experiments combine the factors at each level, so that the total number of combinations (i.e. conditions) studied must be the product of the number of levels of each factor. In the security screening experiment, we collected data on the three posture levels for each of the 12 participants, giving 36 combinations. Note that we did not use each run order for each screener, but balanced the three sets of images across participants so that each participant saw a different image set in their three postures.

## MAIN EFFECTS

For each experiment, there is a structural model, written as in the example earlier ( $X_{ij} = \mu + A_i + B_j + \varepsilon_{ij}$ ). This example is a very simple model where the separate effects of A and B are strictly additive. That means that the effect of Factor A is the same at all levels of Factor B. This is called a ‘main effect’. If we plotted a graph of our measure against the level of Factor A, then there would be parallel graphs for each level of Factor B.

## INTERACTIONS

Experimental results are not usually limited to main effects. The effect of one factor may well depend on the level of another factor: the graphs need not be parallel. To take an example from physics, the combined gas law relates the pressure ( $P$ ), the temperature ( $T$ ) and the volume ( $V$ ) of an ideal gas by

$$PV = kT$$

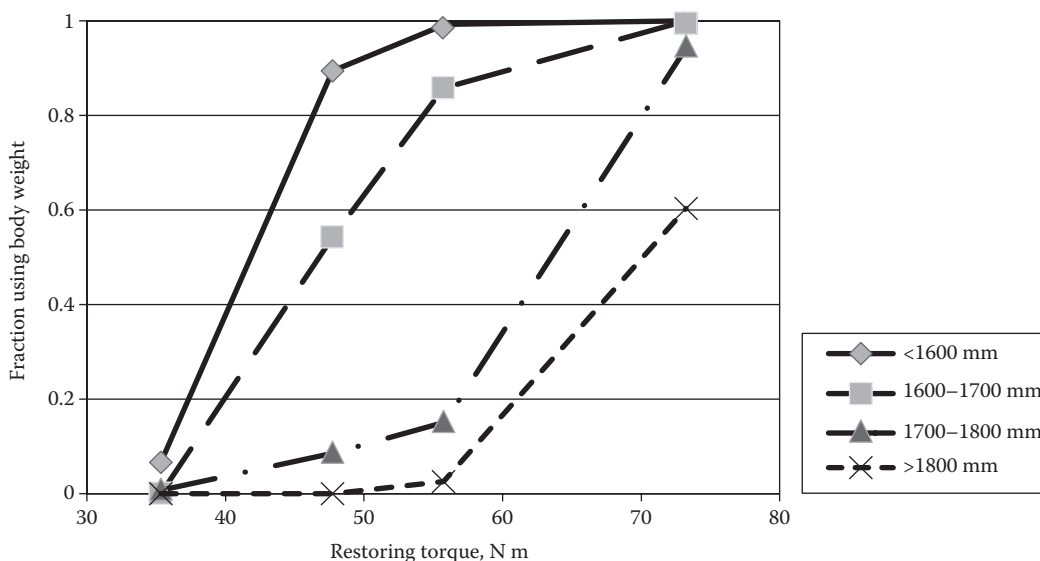
or

$$T = \frac{PV}{k}$$

Thus, the effects of  $P$  and  $V$  on  $T$  are not additive but multiplicative. If we plotted  $T$  against  $P$  at different values of  $V$ , we would get lines that were converging rather than parallel. This is a simple example where it is just a different operator relating the two factors  $P$  and  $V$ . In more complex situations, there may be different joint effects of the two factors. All non-additive combinations of factors are called interactions (which can be described as the effect of one independent variable on the effect of another independent variable on the dependent or measured variable), and are one of the main reasons for performing multifactorial experiments.

An example from the doors study mentioned at the beginning of the chapter is shown in Figure 2.1. It is obvious that the lines joining the data points at each level of participant stature are not parallel, although they do form a pattern that should be familiar to those with E/HF training. Compared to shorter people, taller (and presumably stronger) individuals do not need to use their body weight to open doors until a much higher level of door restoring torque.

In the design and analysis of multifactorial experiments, interactions can be treated statistically by including an extra term in the structural model to represent the combined effect of



**FIGURE 2.1** Example of an interaction from Chang and Drury (2007). The lines reflect four groups of people having different stature.

two factors that is not predictable from their individual additive effects. This is the term  $AB_{ij}$  in the following equations:

$$X_{ij} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{ij}$$

$$\text{Var}(X_{ij}) = \text{Var}(A_i) + \text{Var}(B_j) + \text{Var}(AB_{ij}) + \text{Var}(\varepsilon_{ij})$$

Note that the structural model is still additive, so the variances are also still additive. The additional interaction term can now be tested for significance like all of the other terms against its appropriate error variance. Note also that we need some independent measure of the error variance. This is typically accomplished by repeating measurements under nominally identical conditions. In the doors study, multiple people were observed in each of the 16 combinations, but because the measure was only use/non-use of body weight (a nominal measure), it could not be easily used to estimate error variance. In the case of all three example experiments in this chapter, other assumptions were made in order to obtain estimates of error variance. The safest way is always direct replication, but this may not be possible for all experiments. For example, in the weather displays study also mentioned earlier in this chapter, the same weather front conditions could not be replicated so that only a single measure was possible for each combination of factor levels.

We can obviously extend the multifactorial idea to experiments with more than two factors. With each additional factor, we can find more interactions. For two factors, we can measure the effects of

$$A, B, A \times B.$$

When we add a third factor  $C$ , this becomes

$$A, B, A \times B, C, A \times C, B \times C, A \times B \times C.$$

Interactions are of great importance in E/HF, often because they represent multiple sources of stress on the human in a system, and each additional source may cause a more-than-additive effect as the limits of human capability are reached. Figure 2.1 is an example of this: at the lowest restoring torque, door opening is easy for all of the population, but as restoring torque increases, the impact is greater among the less strong members. Interactions also have importance in the theoretical underpinnings of our discipline. For example, in time-sharing between two tasks, the lack of an interaction implies that the two tasks must be processed serially, while the existence of an interaction implies that some parallel processing is possible (Wickens and Carswell, 2012). Also in visual search tasks, the rate of increase of reaction time with increasing background complexity depends upon the difficulty of discriminating a target from its background. Thus, if target/background discriminability is one factor and number of non-targets in the search field is another, there will be an interaction between these factors on reaction time (Treisman, 1986). In the extreme, if one target is pre-attentive, then the number of non-targets will have no effect on reaction time for that target, leading to an extreme interaction effect. Finding interactions, and finding situations where there are no interactions, are primary objectives of practical E/HF and can form the basis for DoE principles.

## DESIGN OF MULTIFACTORIAL EXPERIMENTS

The basic design of a multifactorial experiment, known as a complete factorial, is to test all combinations of levels of all factors. We just add factors such as  $D$ ,  $E$ ,  $F$ , etc. beyond the aforementioned  $A$ ,  $B$  and  $C$ . This was used in the doors study, where 4 levels of restoring torque were combined with 4 levels of participant stature to give the 16 conditions plotted in Figure 2.1. Also, in the weather display study, 6 pilots were combined with 6 weather front distance and 3 displays to give 108 conditions. In contrast, the security screening study did not use all combinations of 3 postures, 12 participants, 3 run orders and 3 image sets, using only a single image set for each run order. This gave a design comprising 3 postures  $\times$  12 participants and so could only find main effects of posture, participant and run order.

**TABLE 2.1**  
**Design Tableau of a  $3 \times 3 \times 2$  Complete Crossed Factorial Experiment with Three Replications**

| Factor A | Factor B | Participant | Replication 1 | Replication 2 | Replication 3 |
|----------|----------|-------------|---------------|---------------|---------------|
| Level 1  | Level 1  | P1          |               |               |               |
| Level 1  | Level 2  | P1          |               |               |               |
| Level 1  | Level 1  | P2          |               |               |               |
| Level 1  | Level 2  | P2          |               |               |               |
| Level 1  | Level 1  | P3          |               |               |               |
| Level 1  | Level 2  | P3          |               |               |               |
| Level 2  | Level 1  | P1          |               |               |               |
| Level 2  | Level 2  | P1          |               |               |               |
| Level 2  | Level 1  | P2          |               |               |               |
| Level 2  | Level 2  | P2          |               |               |               |
| Level 2  | Level 1  | P3          |               |               |               |
| Level 2  | Level 2  | P3          |               |               |               |
| Level 3  | Level 1  | P1          |               |               |               |
| Level 3  | Level 2  | P1          |               |               |               |
| Level 3  | Level 1  | P2          |               |               |               |
| Level 3  | Level 2  | P2          |               |               |               |
| Level 3  | Level 1  | P3          |               |               |               |
| Level 3  | Level 2  | P3          |               |               |               |

In E/HF, the factor of participant can require special treatment. An example of a complete factorial experiment, with 3 levels of Factor A, 2 of Factor B, 3 of participant and 3 replications (i.e. repetitions) of each combination, is given in Table 2.1. This of course requires a minimum of  $3 \times 3 \times 2 \times 3 = 54$  measurements.

This design can be used to calculate the variance component of all three main effects (A, B, P), all two-way interactions ( $A \times B$ ,  $A \times P$ ,  $B \times P$ ) and the three-way interaction ( $A \times B \times P$ ) as well as a true error variance. Note that each participant is tested at all three levels of Factor A and both levels of Factor B. This is known in E/HF as a ‘within participants’ (also known as repeated measures) design, as both factors are tested on the same participants (P1, P2 and P3). It is clearly a ‘good’ design if there are large inter-participant differences as the ANOVA can calculate the effects of A, B and  $A \times B$  independent of participant. In other words, differences between participants do not contribute to the error variance associated with A, B and  $A \times B$  effects. But what if we cannot test each participant more than once? An example is in comparison of learning technologies where a participant can only learn the task once.

So far, we have only considered designs where all combinations of levels were tested, known in DoE as crossed designs. There is another class of designs for complete factorial experiments known as nested designs, or in E/HF as ‘between participants’ (or independent samples) designs. In these designs, one factor (participant) is nested under other factors so that different participants are tested under different conditions. Table 2.2 shows a design equivalent to Table 2.1, but with participant nested under both Factor A and Factor B so that the 18 different combinations of  $A \times B$  are tested with different participants rather than re-using the original 3 participants of the crossed design in Table 2.1. Such designs have the advantage of minimising confounding due to fatigue or learning (order effects), but differences between participants groups contribute to the error variance, making it more difficult to identify A, B and  $A \times B$  effects.

## WITHIN-PARTICIPANT VERSUS BETWEEN-PARTICIPANT DESIGNS

The main criterion for choice between these two design structures is whether or not the participants change during the course of the experiment. Clearly a participant’s strength will not change much over



**TABLE 2.2****Design Tableau of a  $3 \times 3 \times 2$  Complete Nested Factorial with Three Replications**

| Factor A | Factor B | Participant | Replication 1 | Replication 2 | Replication 3 |
|----------|----------|-------------|---------------|---------------|---------------|
| Level 1  | Level 1  | P1          |               |               |               |
| Level 1  | Level 2  | P2          |               |               |               |
| Level 1  | Level 1  | P3          |               |               |               |
| Level 1  | Level 2  | P4          |               |               |               |
| Level 1  | Level 1  | P5          |               |               |               |
| Level 1  | Level 2  | P6          |               |               |               |
| Level 2  | Level 1  | P7          |               |               |               |
| Level 2  | Level 2  | P8          |               |               |               |
| Level 2  | Level 1  | P9          |               |               |               |
| Level 2  | Level 2  | P10         |               |               |               |
| Level 2  | Level 1  | P11         |               |               |               |
| Level 2  | Level 2  | P12         |               |               |               |
| Level 3  | Level 1  | P13         |               |               |               |
| Level 3  | Level 2  | P14         |               |               |               |
| Level 3  | Level 1  | P15         |               |               |               |
| Level 3  | Level 2  | P16         |               |               |               |
| Level 3  | Level 1  | P17         |               |               |               |
| Level 3  | Level 2  | P18         |               |               |               |

any experiment of reasonable duration, while the same participant's task knowledge and skill will surely change unless very experienced participants are chosen and tested using familiar conditions. Thus, in the doors study, we could have re-used participants, although we did not. Also in the weather display study, highly experienced pilots were used, so that changes between multiple trials on how to react to weather fronts would be unlikely to change, so again, the same participants can be re-used in a crossed design. However, in the security screening study, we found a definite learning across trials on the image sets used, so that even our experienced screeners did change over time. We used a partially within-participants design, but were able to measure and remove any change (learning) effects from our comparison of posture effects. To give other examples where different designs were used, consider the following

Laughery and Drury (1979) used a between-participants design in a study of optimisation skills because it was suspected that techniques learned during the solution of one type of optimisation problem might transfer in an inconsistent manner to other problems, with an adverse effect on bias and variability. Thus five participants were used in each condition, which meant that any comparison between conditions had to be made against between-participant variability. The groups were kept reasonably homogeneous (engineering students) but this in turn limits the generalisability of the results.

Drury et al. (1989a,b) studied the biomechanics and physiology of handle positions on boxes used ten participants, each performing a box holding task using ten handle positions. The within-participants design eliminated the influence of individual differences on the effects, allowing the effects of handle positions on boxes to be detected despite the limited sample size.

No changes to the participant were expected during the box holding experiment, but changes were expected in adaptation. Change occurs in humans in the short term as they fatigue and in the long term as they adapt or learn. With appropriate rest periods, no fatigue was expected (or found) in the box holding task and certainly an hour or two of experimentation on a well-practiced task is unlikely to change either a participant's body strength (adaptation) or box holding technique (learning). Hence, a biomechanical and physiologically limited task is unlikely to exhibit what Poulton's famous (1974) paper called asymmetrical transfer effects. The same cannot be said for



most intellectual skills. What you learn in first solving one optimisation problem is quite likely to affect your performance in solving the next. The transfer can be positive, if the same solution techniques are useful in both problems, or negative, if the solution to the first problem is inappropriate in solving the second. An optimisation task is a priori likely to be closer to an intellectual task than to a biomechanical one, hence the choice of a between-participants design. Any human functions, even anatomical ones, will adapt or change given sufficient time, but the key question is not whether or not change will occur but whether enough will occur to bias the experimental comparison. We can minimise change during an experiment by choosing participants already highly skilled, but as noted earlier that worked for pilots but not for security screeners. We can also provide extensive training in the task so that the typical negative exponential or fractional power law learning curve reaches enough of an asymptote to prevent further changes in task performance. Such techniques would allow the greater power of a within-participants experiment (less variance in comparisons). Extensive task training also helps when the participant pool is limited (astronauts) or non-existent (operators of an entirely novel system). Finally, we may be interested in the response of each individual participant rather than the overall distribution of performance, so that a within-participant design must be used.

There is no reason that an experiment must be entirely a crossed design or entirely a nested design. We can have useful designs that are partially crossed and partially nested, called mixed model designs. However, wherever different conditions are tested on the same participant in a factorial study, we must use a form of analysis called repeated measures ANOVA to correctly capture the contributions of the independent variables and interactions on the variability of the measurements.

The final word on between- vs. within-participant designs is that a between-participant design is always the safer alternative, but may not be practical within resource constraints. They are also subject to the risk that differences between individuals may mask influences of factors, particularly when the number of subjects within each condition is small. Within-participant designs need steps to ensure absence of carry-over effects and a different type of ANOVA but, if designed carefully, help to manage the potentially confounding effects of individual differences on the measurements.

### SAMPLE SIZE, EFFECT SIZE AND POWER

We have seen that the significance level chosen for any statistical test determines the probability of the test giving a false alarm, that is concluding that an effect exists when it truly does not (Type I error). As any E/HF professional can guess, it is not possible to discuss false alarms without also discussing the complementary error, that of failing to conclude that an effect exists when it truly does (Type II error). We have two hypotheses, null and alternative, so that if

$$\text{Significance level} = p(\text{conclude Alternative} \mid \text{Null is true}),$$

then its complement, power, is defined as

$$\text{Power} = p(\text{conclude Alternative} \mid \text{Alternative is true}).$$

Whereas we usually look for a very *low* significance level (0.10, 0.05, 0.01, 0.001), we would like a *high* value of power (0.90, 0.95 etc.).

There are four inter-related factors that need to be considered: significance level, power, effect size and sample size. The effect size is the magnitude of the difference between two means, or the variance between several means. The larger the effect size we are looking for, the easier the statistical testing is, so that for a large effect size, we can have both a low significance level and a high power. It is possible to manipulate the anticipated effect size with DoE. If, for example, age is related to a particular performance measure, a more powerful experimental design would be to

compare groups that are dramatically different in terms of age, as opposed to groups that are closer in age. Because a test statistic is the ratio of an effect size to its standard error, we can also reduce the size of the standard error for any given effect size by taking more samples. Recall that the standard error is calculated by dividing by the square root of the sample size, and it becomes obvious that increasing the sample size is quite an inefficient way to increase power.

In designing our experiment, we have so far not mentioned sample size, except when we have given examples of numbers of participants or numbers of replications per condition. But both of these experimental parameters must be chosen before we can proceed, and we can only logically do this on the basis of the other three underlying variables. If we need to know sample size, we must first decide on effect size, significance level and power of the test. The usual statistical textbook advice is to work with the study commissioner to find values of these parameters, but this is more taxing than asking a client for a significance level alone. One of the authors recently had to produce tableaus of effect sizes and sample sizes for an aviation security study so that the client could make better informed decisions about how to set up an experiment. Such an approach is facilitated by web-based software (e.g. <http://homepage.stat.uiowa.edu/~rlenth/Power/>) and advocated in related journal publications (Lenth, 2001).

The mechanics of calculating sample size given the other three parameters are not simple beyond the ‘toy’ designs of comparing two samples with no other sources of variation, but help is provided in many statistical packages. In addition to the web-based application noted earlier, one author has used the PASS (power analysis and sample size) software to determine sample sizes for experimental designs to be analysed by ANOVA. A much-used package (MINITAB) performs power/sample size calculations for many ANOVAs and some non-parametric statistical tests (e.g. chi-square) but not for complex mixed models. But in the airport security study described at the beginning of this chapter, only nominal data could be collected (e.g. number of threats found, number of false alarms) so that contingency tables and the chi-square test took the place of ANOVAs. The sample sizes were computed manually, although it was later found that they were part of the MINITAB statistical package.

## FRACTIONAL FACTORIAL DESIGNS

The complete factorial design is a powerful and often-used tool in E/HF, but also a costly one. It can find important interactions, but if interactions are known (or assumed) not to exist, it is wasteful of resources. There are special designs that trade off knowledge of interactions for reduced experiment size. We have already seen this in the security screening study where the information on the posture  $\times$  participant interaction was sacrificed to allow a study that would fit the available resources. More formal methods are available, known as fractional factorial designs. As their name implies, they only test a fraction of the combinations used in a complete factorial design.

The simplest fractional factorial for E/HF is probably the Latin Square design. It uses three factors, all at the same number of levels, and counterbalances their appearance so that only  $n^2$  instead of  $n^3$  combinations are tested. The Latin Square ensures that each combination of the levels of each factor occurs once and only once in the  $n \times n$  tableau of the design. Clearly something must be lost for such a large saving in effort and indeed it is. All interactions are sacrificed so that only the three main effects can be calculated and the error variance term is merely the left-over variance when the three main effects are calculated. Even the term ‘sacrificed’ does not capture the whole loss: any interactions are confounded in a complex way with the main effects. A Latin Square should only be used where no interactions are expected or the interactions are known from prior research not to exist. This makes a long list of assumptions that should really be attached to the experimental conclusions, although they rarely are. The main use of Latin Squares in E/HF is in presentation order of different conditions to participants in a within-participant design. Thus, if we have six levels of the Factor (A, B, C, D, E, F) and six participants (1, 2, 3, 4, 5, 6), we can present them in the following trial order (Table 2.3):

In this Latin Square, each level of the factor (usually called a treatment) follows a different order for each participant, helping minimise the effects of unwanted transfer between treatments.

**TABLE 2.3**  
**Example of a 6 × 6 Latin Square Design to Eliminate the Potentially**  
**Confounding Effects of Trial Order**

|               | Trial 1 | Trial 2 | Trial 3 | Trial 4 | Trial 5 | Trial 6 |
|---------------|---------|---------|---------|---------|---------|---------|
| Participant 1 | A       | B       | F       | C       | E       | D       |
| Participant 2 | B       | C       | A       | D       | F       | E       |
| Participant 3 | C       | D       | B       | E       | A       | F       |
| Participant 4 | D       | E       | C       | F       | B       | A       |
| Participant 5 | E       | F       | D       | A       | C       | B       |
| Participant 6 | F       | A       | E       | B       | D       | C       |

(The same pattern can be used for any even number of treatments and participants.) This use of a Latin Square is an excellent alternative to randomisation of treatment order across participants.

More complex fractional factorials have been advocated and used in industrial experimentation, often as part of a design with all factors at 2 levels, called  $2^n$  designs. Like Latin Squares, they only require a fraction of the conditions to be tested, and similarly, they do not allow all interactions to be calculated. They are typically advocated in studying the response of a multivariate industrial process to a selection of the postulated variables in a very economical manner so that the most important variables can be determined efficiently. So far they have not been widely used in E/HF, although a few examples exist (e.g. Bishu et al., 1992; Lin and Radwin, 1998; Naugraiya and Drury, 2009). Rather than  $2^n$  designs, these designs are called  $2^{n-k}$  designs, where  $n$  factors are tested within the resources of an  $(n - k)$  design. The assumption behind fractional factorial designs of all types (see, e.g. Taguchi, 1965) is that higher-order interactions are inherently unlikely. Thus, we can confound these interactions with each other and not calculate them. Each fractional factorial design has a design operator, which is an identity equation showing which effects are confounded with which other effects. Typically, we use a design operator that confounds the main effects and lower-order interactions of interest with (unlikely) higher-order interactions.

The Naugraiya and Drury (2009) experiment was intended as a screening experiment to examine the significance of a large number of factors and interactions for a simulated process control task. It used a  $2^{6-1}$  fractional factorial and examined six factors, each having two levels, with 32 cells rather than the 64 cells required for a full  $2^6$  factorial. Participants were assigned randomly with one to each of the 32 unique conditions tested, a most unusual procedure in E/HF experimental design, but one quite common in industrial experimentation. Each participant performed the simulation four times, ‘producing’ and ‘shipping’ 200 industrial parts under different quality challenges.

The six factors were

- Operator expertise (E)
- Operator training (T)
- Process capability (Cp)
- Challenge direction (D)
- Challenge amount (A)
- Cost criterion (C)

The design operator for the experiment was  $1 = \text{ETCpDAC}$ . This ensures that main effects and low-order interactions are only confounded with the more unlikely higher-order interactions:

- Main effects confounded with five-way interactions
- Two-way interactions confounded with four-way interactions
- Three-way interactions confounded with three-way interactions

Thus, we assume that main effects and two-way interactions are in practice unconfounded. We could try to estimate three-way interactions, but each is confounded with the interaction between the remaining factors (e.g.  $E \times T \times C_p$  is confounded with  $D \times A \times C$ ) so that we cannot disentangle their separate effects.

With six main effects and 15 two-way interactions, much was learned about the task. Note however that the unusual experimental design had only a single degree of freedom for each effect tested, with 103 degrees of freedom for the error variance. Note also that there could be no 'between participants' effect calculated with a single participant per condition.

## SEQUENTIAL EXPERIMENTATION

There is an alternative to fractional factorial experiments that can be useful in E/HF where a sequence of studies is performed instead of a single study. It has the potential to measure which interactions are important rather than merely assuming them away. Also it can deal with factors at more than the two levels assumed in most DoE texts for factorial experiments. With only two levels tested for each factor, we cannot find out much about the underlying response surface (i.e. the shape of the relationships between the independent variables and the effects). The strategy is to perform a  $2^n$  complete factorial, measure which interactions are important, then perform a set of experiments with the desired levels of each factor, but only for factor combinations that have measurable interactions. As an example, suppose the design we would like to run has five factors at the following levels:

A at 5 levels  
B at 2 levels  
C at 3 levels  
D at 3 levels  
Replications at 2 levels to provide an error estimate

The full factorial will need  $5 \times 2 \times 3 \times 3 \times 2 = 180$  trials. It will allow calculation of

4 main effects: A, B, C, D  
6 two-way interactions: AB, AC, AD, BC, BD, CD  
4 three-way interactions: ABC, ABD, ACD, BCD  
1 four-way interaction: ABCD  
1 error term

Perhaps, we do not need all of these from a single grand design. We can use a  $2^4$  complete factorial with 2 replications as a screening experiment specifically to test for interactions. This will require  $2 \times 2 \times 2 \times 2 \times 2 = 32$  trials. Then run experiments on the significant interactions. For example if only  $A \times B$  and  $C \times D$  are significant, then we can run two additional smaller experiments:

$A \times B$  with 5 levels of A, 2 levels of B and 2 replications, requiring  $5 \times 2 \times 2 = 20$  trails  
 $C \times D$  with 3 levels of C, 3 levels of D and 2 replications, requiring  $3 \times 3 \times 2 = 12$  trails

We can thus measure all of the effects we were initially interested in with  $32 + 20 + 12 = 70$  trails instead of the original 180 trials. Of course, not all screening experiments produce the same interaction structure, so we could go all the way from every interaction being significant (requiring 180 trials) to no interactions being significant (requiring only  $(5 + 2 + 3 + 3) \times 2$  replications = 26 trails). We can always re-use data between experiments if that is logically possible, for example the 10 combinations in the  $A \times B$  experiment include  $2 \times 2 = 4$  conditions that have already been studied, leaving only 6 additional conditions with 2 replications each.

## ANALYSIS ALTERNATIVES

Throughout this chapter, we have assumed that ANOVA will be the analysis method, primarily because it is well suited to multifactorial experimentation, and almost any E/HF study will have to be multifactorial. ANOVA uses the additive property of variances to decompose a total experimental variance into components associated with each variable and interaction. In simple cases, such as complete factorial designs with replications, we can have independent tests of each factor and interaction. With more complex, but less complete designs, such as fractional factorials or Latin Squares, we forgo some independence for experimental convenience or even for study feasibility. The familiar ANOVA table can be found in any DoE text and will not be repeated here. There are strict rules for how to compute the significance of effects based on their F statistic value, depending upon the structural model of the experiment. This tells the components of variance and thus what denominator to use in the F test.

Most E/HF professionals will not calculate variance components and F-values by hand, relying on statistical packages to perform the computations. These, such as SPSS (Statistical Package for the Social Sciences) or MINITAB, will require the user to input a structural model to control the computations. That is why understanding ANOVA and models remains important in times of automated computation. Failing to treat repeated measures correctly, confusing fixed and random effects or failing to check the ANOVA assumptions are all analysis errors that can be committed by experienced E/HF professionals. The use of packages where ease of use is a design feature should not reduce the diligence of the experimenter at the analysis phase. In this section, we enumerate various issues with analysis, examining what to do when there are multiple dependent variables, how to test assumptions (and what to do if they are not met), and sources for the many different statistical analysis packages available today.

## DEALING WITH DATA

We have discussed experimental design and ANOVA as if there were a single number or measure in each cell of the design, but we now need to expand beyond this. Each cell is the data from a single replication of each combination of factor levels in the design. Assuming that issues of reliability, validity, etc. have been addressed, the first thing to note is that we do *not* put the data into a spreadsheet and calculate means. Each data point must be kept separate: ANOVA procedures will provide table of means, variances and confidence intervals *ad nauseam*. Keeping data separate provides the full degrees of freedom for error variance, helping to ensure that the power of the tests is maintained. Second, many measures come in the form of a continuous variable recorded over a time interval, such as the record of car position in a lane while driving, or the continuous movement of the centre of gravity of a standing operator. In these cases, measures must be derived from the continuous records, for example root-mean-square error or average position in each dimension. It is this number which becomes the data for the ANOVA. These are quite simple matters to deal with and have usually been addressed during the initial design of the experimental study.

Quite often, however, there will be multiple measures per cell of different aspects of performance or well-being or both (i.e. multiple effects). In the security screening study, we measured four performance variables (hits, false alarms, time per image when a threat was found, time per image where no threat was found) and four well-being variables (two measures of body part discomfort, NASA TLX, number of non-work-related movements), giving 8 dependent variables. The obvious way to proceed is to perform an ANOVA on each dependent variable, but this raises two problems. First, because we are performing 8 tests for each factor or interaction, the likelihood of concluding that a statistically significant relationship between one or more independent variables and an effect variable exists when in fact it does not exist (i.e. the likelihood of Type 1 error) increases. Another way to think about this is that the significance levels become inflated. If we chose a 1 in 20 chance of false alarm for any single variable, then the likelihood of having at least one false alarm would be  $(1 - (1 - 0.05)^8) = 0.33$ , which is not what we had planned. Second, it might be that the dependent variables show similar patterns, so that even if one variable is not significant, the same pattern across several variables

may be significant. The standard way to proceed is to perform a multivariate analysis of variance, or MANOVA, across the complete suite of dependent variables. Then, if any factor or interaction proves significant, univariate ANOVAs can be run to determine which variables were responsible. This procedure does not inflate the significance levels, and provides an orderly exploration of the data.

Another technique that can be a powerful tool for reducing a large suite of data to a more manageable number of orthogonal and hence independent tests is factor analysis. Factor analysis (nothing to do with factors in DoE) groups together dependent variables with high inter-correlations. With a modification called Varimax rotation, it will produce a small number of new dependent variables called, confusingly, factors that are orthogonal to each other and summarise a large fraction of the total variability in the data set. It has a long history in the social sciences, and has been used many times in E/HF, from early papers (e.g. Drury and Daniels, 1980) to more recent studies (Ryan et al., 2009). In the security screening study, it was used to explore the inter-correlation matrix of all 8 dependent variables. We found three factors that met the usual criterion for significance: performance (4 variables), posture (3 variables) and workload (TLX only). Each could be analysed in the confidence that only three independent tests were being carried out, and that all of these new independent variables were orthogonal. To simplify the interpretation of the findings, the ANOVA results associated with each of the individual measures that contributed to each of the orthogonal factors were reported in the paper.

## DEALING WITH ASSUMPTIONS

Various assumptions have been made in this treatment of experimental design, assumptions that can often be tested directly from the data collected. We have already considered the additivity assumption and shown how intersection terms can extend the simple additive model of main effects to the generally more interesting interaction effects. There are other ways to deal with non-additivity in special cases. The combined gas law used as an example earlier has a multiplicative effect of P and V on T. Many human functions are multiplicative, particular sensory functions. In any E/HF study where effects look multiplicative, logarithms are a useful transformation tool to allow ANOVA while still preserving the model structure. It is simple to transform the common gas law into an additive function by taking logarithms of the equation:

$$\text{Ln}(T) = \text{Ln}(p) + \text{Ln}(V) - \text{Ln}(k)$$

Additivity is now satisfied. Transforms are an integral part of many science and engineering formulations, and are frequently used in measuring human performance. Examples are the use of the decibel scale in auditory perception and the same scale to study human tracking behaviour.

Transforms can also help with the ANOVA normality assumption. We can test this assumption by having the analysis package plot residuals (the difference between a data point and its expected value from the ANOVA model) as a cumulative normal distribution. Either visually or statistically, we can determine whether the normal distribution is a good fit to the data. If it is, then the ANOVA is valid for the normality assumption: if not, the pattern of deviations from the normal distribution function provides clues to suitable transforms.

Some measures of human performance and well-being are quite normally distributed, but task completion times are often not. Most performance time data will have a lower bound beyond which the human cannot react or move any more rapidly. But the upper bound is often unlimited, leading to time distributions that are positively skewed, with a longer 'tail' to the right. A lognormal distribution is often a good fit, so that transforming the raw time data to  $\text{Ln}(\text{time})$  will produce normally distributed data suitable for use in ANOVA. In odd cases, even more skewness is expected: Search times in extended search tasks (such as security screening) are expected theoretically to follow a negative exponential distribution (Morawski et al., 1980). A  $\ln(\text{time})$  transform is usually sufficient to normalise the data, however. Most statistical texts include at least something on transformation to improve normality and homogeneity of variance or to meet additivity assumptions, removing interactions (e.g. Winer et al., 1991, pp. 354–358).



Probability or frequency data generated from repeated measures made on nominal or categorical scales, particularly the (0,1) form of data, are inherently non-normal. There is much discussion in the statistical and social science literature of the legitimacy of using ANOVA for categorical data, with some claiming it can be used with minimum danger, while others recommend arc-sine or logistic transforms before using ANOVA. Rather than take sides in this, articles such as Jeager (2008) should be consulted for the most recent findings.

This introduction of nominal data brings in the whole question of alternative forms of data analysis to ANOVA itself. Nominal data can often be best analysed using contingency tables and the chi-square or Fisher's exact tests. As texts on categorical data analysis (e.g. Agresti, 1996) explain, contingency tables can go far beyond the  $2 \times 2$  example given in most statistics texts. One-way, two-way and even three-way tables can be analysed rather simply to give the equivalent of ANOVA for nominal data. One can even calculate standardised residuals to provide post hoc comparisons and determine which cells in the design have significantly high contributions to the overall chi-square statistic. The use of chi-square for contingency table analysis is particularly suitable for the relative small counts of rare events during experiments, such as errors. For many years (e.g. Drury and Daniels, 1975), one author has used ANOVA to analyse task completion times while using chi-square to analyse error frequencies.

The final assumption in ANOVA is homogeneity of variance. The homogeneity of variance assumption requires the variability of the data to be the same across the conditions tested with ANOVA. This is treated in most texts (e.g. Winer, 2012), and tests such as Bartlett's test and Scheffe's test are recommended to be run on the data to check that the variance is not different across cells in the design. If the test is not satisfied, the analysis can sometimes proceed with a transformation. For example if the variance in each cell is related to the cell mean, then a logarithmic or square-root transform will help homogeneity.

## DEALING WITH ANALYSIS PACKAGES

One recent survey found over 40 available software packages for DoE and statistical analysis of the data. Online comparisons between features and costs are easily located using Internet search engines. As already noted, people do not perform ANOVAs manually any longer, although most E/HF practitioners learn this skill in elementary statistics courses. With the advent of data-based quality philosophies in industry (e.g. Six-Sigma) has come the wide demand for statistical analysis tools. These quality philosophies rarely take into account E/HF topics, so it is unsurprising that some statistical packages also omit the unique issues of experimenting with people, for example repeated measures designs. The range of software is huge, as is the range of costs. Some excellent software is free, such as 'R' which is integrated in the Internet-based *e-Handbook of Statistical Methods* available at the *National Institute of Standards and Technology* (<http://www.itl.nist.gov/div898/handbook/index.htm>) in the United States. Others cost many thousands of pounds/dollars, although academic licences are usually available to ease the cost to students. Some have free trial introductions, enabling users to test their suitability using their own data. Most packages work on the usual computer platforms of Windows, Mac-OS and Linux, although MINITAB and SAS/Stat appear (2014) to have stopped Mac-OS support. The authors have used MINITAB, SPSS, SAS, Number Cruncher Statistical System (NCSS) and PASS at different times, and so these are the basis for the following comments. Note that statistical software is updated frequently, so that the lack of a particular feature in one package may not be forever.

Three general observations are in order when considering statistical packages. First, the experimenter should be extremely cautious when attempting to use general-purpose spreadsheets (e.g. Excel) for statistical testing of experimental data. While such programs have statistical analysis routines, they are very simple and may not have much utility in E/HF experimentation. They may also make assumptions (e.g. on F-tests) that the experimenter may not have intended, and rarely have the ability to perform ANOVA with the appropriate main effect and interaction structure discussed previously in this chapter. Second, when using a statistical analysis software package, the experimenter should always perform a single analysis on the complete design (i.e. the full model), rather than

multiple one-way ANOVAs on each factor. While statistics texts never advocate the use of multiple one-way ANOVAs instead of using the full model, unfortunately, students or industrial users who have only had brief introductions to statistics seem more comfortable using such an approach. The complete ANOVA will calculate all available effects, and thus remove their variance contribution from the residual error, leading to much more powerful F-tests. Lastly, it is recommended that the experimenter, after designing the experiment, generate random data to populate the data array *before* running the experiment. In this way, the experimenter can see what the output will be like, which tests are possible and whether any factors are confounded. Running the analysis on random data first is a good final check on the design before data collection.

All of the packages mentioned perform ANOVAs to considerable levels of complexity, except for PASS which is a dedicated power and sample size calculation tool. All have moved many years ago from a command line interface to a graphical user interface. Data can be entered directly into a spreadsheet-style tableau, or pasted in from a file collected as part of the experiment. Import and export of data files between statistical packages is an important asset, allowing the strengths of different packages to be complementary. Most packages will accept standard spreadsheet files, such as .xls files from Excel, so that this intermediary can often be used if no automatic data transfer routine is available.

MINITAB grew from a very simple and small data analysis system, as its name implies. It is now a very complete package aimed at industrial users, with much support for process analysis and statistical process control. It works well for fractional factorial designs, both generating them and analysing the results. At times, the format of the structural model can be confusing in MINITAB, but it generally will cover your experiment if the model is specified correctly. NCSS has also grown considerably since it was first introduced as software for the computer industry. It includes complex ANOVAs, with options such as repeated measures designs useful to E/HF professionals. SAS, specifically SAS/STAT, is the statistical analysis component of a much broader range of business analytics. SPSS (Statistical Package for the Social Sciences), now part of a suite of statistical software from IBM, is more oriented to the social sciences as its name implies, with good support for ANOVA models where participants are a factor. It, like MINITAB, SAS/STAT and NCSS, will perform ANOVAs and MANOVAs, calculate components of variance and effect sizes, check for normality and homogeneity of variance and provide a variety of post-hoc comparisons between specific levels of a factor or interaction. In addition, all provide multivariate analysis such as factor analysis, non-parametric statistics and analysis of contingency tables.

This chapter has framed the traditional benefits and challenges of statistical DoE in terms of the work of the E/HF professional. Whether the issue is our basic science or more applied knowledge, the same principles apply: choice of dependent and independent variables, choice of sampling procedures and sample size, detailed design of the experiment to maximise effectiveness without excessive time and cost and the ethics of experimentation on humans. Without an experiment, it is difficult to impute causality to results. Without a designed experiment, it is difficult to explore the many causal factors known (or suspected) to influence human well-being and system performance. While excellent statistical texts provide many details of how to design and analyse experiments, they must use very general examples to make them applicable to a broad audience. In our discipline, we can use our specialist knowledge and insights to make more informed choices of what to vary, what to measure and what to control so as to remove sources of contamination or uncertainty. Experimental design is not performed in isolation: it is one more powerful means for us to make decisions about future actions.

## REFERENCES

- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*, Wiley, New York.
- Bishu, R.R., Wei Wang, Hallbeck, M.S. and Cochran, D.J. (1992). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 36, 816–820. Santa Monica, CA: Human Factors and Ergonomics Society.
- Chang, S.-K. and Drury, C.G. (2007). Task demands and human capabilities in door use. *Applied Ergonomics*, 38(3), 325–335.



- Drury, C.G. (1994a). Function allocation in manufacturing. In: S.A. Robertson (Ed.), *Contemporary Ergonomics 1994*, Keynote address to the Ergonomics Society Meeting, Taylor & Francis Group, London, U.K., pp. 2–16.
- Drury, C.G. (1994b). The speed-accuracy trade-off in industry. *Ergonomics*, 37, 747–763.
- Drury, C.G., Cardwell, M.C. and Easterby, R.S. (1974). Effects of depth perception on performance of simulated materials handling task. *Ergonomics*, 17, 677–690.
- Drury, C.G. and Daniels, E.B. (1975). Performance limitations in laterally constrained movements. *Ergonomics*, 18, 389–395.
- Drury, C.G. and Daniels, E.B. (1980). Predicting bicycle riding performance under controlled conditions. *Journal of Safety Research*, 12(2), 86–95.
- Drury, C.G. and Dawson, P. (1974). Human factors limitations in fork-lift truck performance. *Ergonomics*, 17, 447–456.
- Drury, C.G., Deeb, J.M., Hartman, B., Woolley, S., Drury, C.E. and Gallagher, S. (1989a). Symmetric and asymmetric manual materials handling. Part 1: Physiology & psychophysics. *Ergonomics*, 32(5), 467–489.
- Drury, C.G., Deeb, J.M., Hartman, B., Woolley, S., Drury, C.E. and Gallagher, S. (1989b). Symmetric and asymmetric manual materials handling. Part 2: Biomechanics. *Ergonomics*, 32(6), 565–583.
- Drury, C.G. and Forsman, D.R. (1996). Measurement of the speed accuracy operating characteristic for visual search. *Ergonomics*, 39, 41–45.
- Drury, C. G. (1997). Ergonomics Society Lecture 1996: Ergonomics and the Quality Movement. *Ergonomics*, 40(3), 249–264.
- Drury, C.G., Holness, K., Ghylin, K.M. and Green, B.D. (September 2009). Using individual differences to build a common core dataset for aviation security studies. *Theoretical Issues in Ergonomics Science*, 10(5), 459–479.
- Drury, C.G., Hsiao, Y.L., Joseph, C., Joshi, S., Lapp, J. and Pennathur, P.R. (2008). Posture and performance: Sitting vs. standing for security screening. *Ergonomics*, 51(3), 290–307.
- Drury, C.G., Mirka, G.A. and Marras, W.S. (2007). *The Publishable Practicum, Contemporary Ergonomics 2007*, Taylor & Francis Group, London, U.K., pp. 247–252.
- Drury, C.G. and Sinclair, M.A. (1983). Human and machine performance in an inspection task. *Human Factors*, 25, 391–399.
- Fisher, D.L., Schweickert, R. and Drury, C.G. (2012). Mathematical models in engineering psychology: Optimizing performance. In: G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics*, John Wiley & Sons, Inc., Hoboken, NJ, pp. 962–989.
- Hoffmann, E.R. (1992). Fitts' law with transmission delay. *Ergonomics*, 35(1), 37–48.
- International Ergonomics Association, <http://www.iea.cc/>.
- Jackson, D. (1956). A short form of the Witkin's embedded-figures test. *Journal of Abnormal and Social Psychology*, 53(2), 254–255.
- Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
- Java applets for power and sample size, <http://homepage.stat.uiowa.edu/~rlenth/Power/>.
- Laughery, K.R. and Drury, C.G. (1979). Human Performance as Strategy is a Two-Variable Optimization Task. *Ergonomics*, 22(12), 1325–1336.
- Latorella, K.A. and Chamberlain, J.P. (2001). Decision-making in flight with different convective weather information sources: Preliminary results. *Focusing Attention on Aviation Safety. Proceedings of the 11th International Symposium on Aviation Psychology*, Columbus, Ohio.
- Lin, M.L. and Radwin, R.G. (1998). Agreement between a frequency-Weighted filter for continuous biomechanical measurements of repetitive wrist flexion against a load and published psychophysical data. *Ergonomics*, 41(4), 459–475.
- Lenth, R.V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55, 187–193.
- Morawski, T., Drury, C.G. and Karwan, M.H. (1980). Predicting search performance for multiple targets. *Human Factors*, 22(6), 707–718.
- Mozzall, J. and Drury, C.G. (1996). Effects of physical exertion on task performance in modern manufacturing: A taxonomy, a review and a model. *Ergonomics*, 39, 1179–1213.
- Naugraiya, M. and Drury, C.G. (2009). A fractional factorial screening experiment to determine factors affecting discrete part process control. *Theoretical Issues in Ergonomics Science*, 10(1), 1–17.
- NIST Statistics Handbook, <http://www.itl.nist.gov/div898/handbook/>.
- Oxford English Dictionary, [www.oxforddictionaries.com](http://www.oxforddictionaries.com).
- Pew, R.W. (1969). The speed-accuracy operating characteristic. *Acta Psychologica*, 30, 16–26.
- Poulton, E.C. (1974). *Tracking Skill and Manual Control*, Academic Press, New York.

- Ryan, B., Wilson, J.R., Sharples, S., Morrisroe, G. and Clarke, T. (2009). Developing a Rail Ergonomics Questionnaire (REQUEST). *Applied Ergonomics*, 40, 216–229.
- Siegel, S. and Castellan, N.J. (1988). *Non-Parametric Statistics for the Behavioural Sciences*, 2nd edn., McGraw-Hill, New York.
- Taguchi, G. (1986). *Introduction to quality engineering: Designing quality into products and processes*. Tokyo, Japan: Asian Productivity Organization.
- Treisman, A. (1986). Properties, parts and objects. In: K.R. Boff, L. Kaufman and J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, Wiley, New York.
- Wickens, C.D. and Carswell, C.M. (2012). Information processing. In: G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics*, John Wiley & Sons, Inc., Hoboken, NJ, pp. 117–161.
- Winer, B.J. (2012). *Statistical Principles in Experimental Design*, Literary Licensing, LLC, New York.
- Winer, B.J., Brown, D.R. and Michels, K.M. (1991). *Statistical Principles in Experimental Design*, McGraw-Hill Inc., New York.
- Witkin, H. (1950). Individual perceptions of ease of perception. *Journal of Personality*, 19(1), 1–15.

---

# 3 Study and Analysis of Complex Cognitive Work

*Ann Bisantz, Emilie M. Roth and Jennifer Watts-Englert*

## CONTENTS

|                                                         |    |
|---------------------------------------------------------|----|
| Introduction and Scope .....                            | 61 |
| Perspectives for Analysing Complex Cognitive Work ..... | 62 |
| Overview of Data Collection Methods .....               | 63 |
| Observational Methods .....                             | 63 |
| Setting .....                                           | 64 |
| Degree of Interaction .....                             | 64 |
| Parallel Sources of Data .....                          | 64 |
| Collecting Observational Data .....                     | 65 |
| Interview and Focus Group Methods .....                 | 66 |
| Generative Methods .....                                | 67 |
| Theoretical Perspectives and Analysis Frameworks .....  | 68 |
| Cognitive Psychology and Cognitive Science .....        | 68 |
| Cognitive Systems Engineering .....                     | 68 |
| Naturalistic Decision-Making .....                      | 69 |
| Ethnography .....                                       | 69 |
| Interaction Design .....                                | 70 |
| Scenario-Based Design .....                             | 71 |
| User Needs/Requirements Analysis .....                  | 71 |
| Outputs and Applications of the Methods .....           | 72 |
| Interfaces for Monitoring and Control .....             | 72 |
| Support throughout the Design Cycle .....               | 73 |
| Human-Centred Automation .....                          | 73 |
| Corporate Knowledge Capture .....                       | 74 |
| Intellectual Property .....                             | 74 |
| Techniques for Making an Impact .....                   | 75 |
| Multidisciplinary Teams .....                           | 75 |
| Representing Study Findings and Implications .....      | 76 |
| Conclusions .....                                       | 76 |
| Acknowledgements .....                                  | 76 |
| References .....                                        | 76 |

## INTRODUCTION AND SCOPE

Work analysis in human factors and ergonomics has shifted from focusing primarily on the physical demands of work to include both cognitive and socio-technical components of human-technology systems. This shift has been motivated by fundamental changes in work due to the advent of information technology and automation as well as high-profile system failures (e.g. Three Mile Island, numerous aviation accidents, and military incidents such as the USS *Vincennes* incident), which

drew attention to the need to understand and support cognitive activities associated with complex system control. With this shift has come an evolution of work analysis methods from the time and motion study methods developed as part of the scientific management movement at the beginning of the twentieth century (Gilbreth and Gilbreth, 1917; Taylor, 1911) to task analytic techniques (e.g. Hierarchical Task Analysis; Annett and Duncan, 1967). These methods allowed the physical, perceptual, and cognitive demands of task components to be compared against human capabilities, compared with those methods which support analysis of complex cognitive and collaborative work (Bisantz and Roth, 2008; Crandall et al., 2006; Rasmussen et al., 1994; Vicente, 1999). In addition, there has been cross-fertilization of methods from the social sciences with the more traditional work analysis techniques and research and design questions addressed in human factors. For example, ethnography is now relatively commonplace in human factors (Blomberg et al., 1993; Hammersley and Atkinson, 1983). This intersection has informed both research and design, particularly regarding studies on the impact and design of modern computer and information technologies (Beyer and Holtzblatt, 1998; Carroll, 1995, 2000; Nardi, 1997; Sachs, 1995; Suchman and Trigg, 1991; Zuboff, 1987).

This chapter focuses on methodologies intended to explicitly identify the requirements of cognitive work in order to anticipate contributors to performance problems (e.g. sources of high workload, contributors to error) and specify ways to improve individual and team performance through, for example, new forms of interfaces, automation schemes, communication and coordination methods, or training approaches. Specifically, the chapter covers the various foci for analysing complex, cognitive work; provides a broad overview of the types of data collection methods typically employed; describes the theoretical perspectives and analytic frameworks that inform the analysis and outputs of the collected data; presents various outputs and applications of the methods; and describes techniques that have been used to ensure that the analyses and applications have an impact on system design and deployment.

Bisantz and Roth (2008) provide a more comprehensive literature review of cognitive analysis methods and applications. Crandall et al. (2006) and Hoffman and Militello (2008) provide in-depth coverage and 'how-to' guidance. Annett (2000) provides an excellent discussion of historical roots and precursors of cognitive analysis methods.

## PERSPECTIVES FOR ANALYSING COMPLEX COGNITIVE WORK

Methods for analysing complex cognitive work typically provide descriptions which are informed by two mutually reinforcing perspectives: a focus on the work and domain-driven factors which shape, support and constrain performance, and a focus on the knowledge, strategies and skills held by domain practitioners (particularly, expert practitioners) that allow them to operate successfully in the domains (Bisantz and Roth, 2008).

The analysis of domain characteristics provides a framework for understanding the goals and constraints present in the domain; the resources available to achieve system goals or purposes; the information necessary to successfully oversee, control and intervene in the work domain; complexities such as interactions, uncertainty and time pressure that make system control challenging; and the likely cognitive demands imposed by these factors. For instance, analyses can identify what information is available to practitioners and whether key resources necessary for control are present, interacting goals or resource limitations that make it difficult to achieve success, information that needs to be sensed to allow operator control and contexts in which automation or other aids could be effectively deployed.

The second, complementary perspective examines the goals, motivations, knowledge, skills and strategies of domain practitioners that enable domain practitioners to operate at an expert level as well as the cognitive factors that limit the performance of less experienced individuals (e.g. incomplete or inaccurate mental models). The results can be used to identify opportunities to improve performance either through training or through the introduction of systems that more effectively support cognitive performance.

Across these perspectives, there is a focus on communication, collaboration and task coordination among multiple human and technological agents. Aspects of both the work domain (e.g. information displays, shared planning systems) and expert performance (e.g. coordination strategies, implicit and explicit communication methods) are important contributors to successful communication and coordination. Often, there is particular interest in understanding the role that existing tools or ‘cognitive artefacts’ play in supporting work, including the support of communication and task coordination. Such tools can range from low-tech (and often flexible) artefacts such as whiteboards and post-it notes; to specially designed, task-relevant objects (e.g. a military patrol reporting form; a pre-filled and labelled drug syringe; a light meter designed for photographers); to high-tech digital technologies such as information displays and messaging systems. For instance, investigations of cognitive performance, expertise and in situ use of technology and artefacts have been used to identify ‘workaround’ strategies that signal the need for more effective cognitive support (e.g. Mumaw et al., 2000; Vicente et al., 2001), as well as effective strategies or artefact properties that should be preserved or reproduced as new technology is introduced (e.g. Bisantz et al., 2010; Roth and Patterson, 2005; Roth et al., 2006; Xiao, 2005).

Across the many different data collection methods and analytic traditions associated with the analysis of complex, cognitive work, researchers and practitioners may choose to place the emphasis of their analysis on either the work domain, or practitioner expertise in response to the goals of the analysis. However, the two perspectives are clearly mutually informing. Task demands interact with practitioner expertise, strategies and practices, and work supports to make aspects of system control more or less challenging. To effectively support system design and performance-aiding efforts, analyses must reveal these complex interdependencies by taking into account both perspectives (Hoffman and Lintern, 2006; Potter et al., 2000).

## OVERVIEW OF DATA COLLECTION METHODS

Numerous data collection techniques have been employed to document and describe these various foci. Due to the combined requirements to understand complexities of the work domain, along with strategies, skills and knowledge deployed by experts, a common thread among these methods is that they rely heavily on the participation of domain experts: as interview respondents, as the focus of observations or as participating members of the design teams. An analysis of complex, cognitive work simply cannot be completed without the involvement, at some level, of those individuals who perform the work. Additionally, the observations and data produced by these methods can be driven by both a top-down, and bottom-up approach. In the former case, theoretical perspectives or specific research questions (e.g. developing a model of naturalistic decision-making in a particular work setting, understanding the role of a specific artefact in communication) will guide what observations are collected. In the latter case, and often in order to generate novel hypotheses or models, observations will be influenced by the situation, opportunistically. Such data can be analysed using a grounded theory approach (Glaser and Strauss, 1967) to develop themes or hypotheses, which can in turn influence more focused data collection.

Commonly used techniques include observations, interviews and focus groups, process-tracing methods, generative methods like diaries and collages, artefact and document analysis, and analysis of critical incidents.

## OBSERVATIONAL METHODS

An important method for gaining insights into complex, cognitive work is the observation of work activities. Observational methods can vary along a number of key dimensions (e.g. see Bisantz and Drury, 2005), many of which are relevant to the use of these methods for understanding complex, cognitive work. These choices include the setting for observations, degree of interaction among the

research and those being observed, whether observations are drawn from real-life or through audio or video-taping sessions and the degree of structure in collecting and reporting observational data.

### **Setting**

While typical applications of observational methods occur in naturalistic settings as work activities are unfolding (e.g. Roth et al., 2004), it is important to recognize the difference between studies in real world settings, often considered field or case studies, and the use of observational, or non-intrusive data-gathering techniques. It is possible to conduct studies in field settings which do not rely only on observation as a measurement technique (e.g. eye gaze patterns; Montague et al., 2011). Likewise, it is possible to make observations in settings which are pseudo-realistic, such as task simulators (Bowers et al., 1998) or training exercises (Artman, 2000), or even in laboratory settings, where observations can be made of behaviour or activities during the performance of more controlled tasks. In some cases, individuals may be instructed to think aloud as they perform the task, to provide an ongoing verbal protocol of the task (Bainbridge and Sanderson, 1995; Ericsson and Simon, 1993). For instance, Gorman and Militello (2004) observed functionally blind users performing specified Internet search tasks in conjunction with a screen reader. Users were asked to think aloud during the task in a laboratory environment, and decision models were developed to describe their activities.

Observational studies in field settings are particularly useful for identifying mismatches between how work is depicted in formal processes and procedures and how it is actually performed, often revealing 'home-grown' tools and workarounds that domain practitioners generate to cope with aspects of task complexity that are not well supported (e.g. Roth et al., 2006). Differences between 'canonical' descriptions of work and actual work practice can reveal opportunities to improve performance through more effective support.

### **Degree of Interaction**

At one extreme, the researcher or analyst acts as an observer, or the instrument of measurement, with minimal interaction with the people and situations being observed. This type of observation may be appropriate in cases where research questions are well defined, where the researcher has a strong understanding of domain activities and where participant actions are easily observed. More typically, however, real-time observations in actual work settings are often combined with informal interviews conducted as the task progresses. In some cases, participant observation methods are employed in which analysts participate in the work performance (often in an 'apprenticeship' capacity). For instance, Burns and Vicente (2000) acted as human-machine design consultants, and thus as participant-observers, during their observations of the interaction between human factors and other design constraints.

### **Parallel Sources of Data**

In most studies that utilize observational techniques, additional measurement techniques are used. Objective records of unfolding events such as events and process variable states may be collected and combined with the observations that are made (either in real time or from recordings) to create a rich protocol or 'process trace' that captures the unfolding events and task activities, thus allowing the activities of operators to be understood within the context of the task itself (Woods, 1993). For instance, Cook and Woods (1996) utilized process-tracing methods, collecting data regarding patient states (e.g. physiological parameters) and conducting concurrent interviews in addition to making observations of practitioners using medical devices. Similarly, Seagull and Sanderson (2001) constructed process-tracing logs which included surgical events and equipment states, as well as observed activities. Observations of artefact use, teamwork and collaboration, and information use may be supplemented with interviews (either concurrently, as the work activity progresses, or in a more formal interview setting), focus groups or through the analysis of archival data sources. In another example, Degani and Wiener (1993, 1997)



supplemented observations of pilot checklist use with interviews with pilots and analysis of archival data (aviation accident and incident records).

### Collecting Observational Data

Observations can vary from unstructured, opportunistic field notes to more structured observations based on predetermined categories. Regardless of the format, observations are guided or shaped by research questions and/or analysts' expertise, goals and theoretical perspectives. Data collection methods can range from decidedly low-tech (pen and paper) to audio and video recordings, to computer-supported data collection. For instance, Patterson and Woods (2001) conducted observations that focused on space shuttle mission control shift change and handovers during an actual space shuttle mission. They combined observations with handwritten logs and spontaneous verbalizations of the controllers (captured via audiotape), along with flight plans, to identify handover activities that were related to fault management, re-planning and maintaining common communicational ground. Handwritten notes can be free form (in the case of unstructured data) or rely on scales and forms developed based on structured categories of data to be collected (Sharples et al., 2011). Specifically developed software programs combined with portable or handheld computers can be used to allow coding of real-time activities. There are also hardware and software systems which facilitate analysis of real-time or videotaped data (e.g. Noldus™, NVivo™, Morae™, also see Mackenzie and Xiao, 2011). These systems provide functionality such as time-based assignment of codes and descriptions to video segments and integrated control of video.

Video and audio recordings can be used to capture observational data for later analysis. Advantages of collecting video data, and making observations from the recordings, are numerous. Unlike in-person observation, situations captured in tapes can be reviewed multiple times, in order to fully assess aspects of complex or fast-paced activities. For example, Kirschenbaum (2004) observed groups of weather forecasters in either their everyday work setting, performing normal forecasting duties, or in a simulated shipboard forecasting centre, working on a provided scenario. Team activities, along with think-aloud protocols, were captured via videotape to allow for detailed qualitative data analysis of cognitive activities related to weather forecasting. Video records also make it possible to collect cued retrospective explanations of task performance by the individuals who participated in the task (Hoc and Leplat, 1983). Seagull and Xiao (2001) used video recordings on which eye-tracking data had been superimposed to study a surgical procedure. The recordings were made from the perspective of the physician performing the procedure (wearing mobile recording and eye-tracking equipment). The eye-tracking data indicated where (in the operating room) the physician looked throughout the procedure. The tapes were reviewed by the physician and other subject matter experts (SMEs) to determine what the physician had to look at to accomplish the task, what that information would indicate and why it was sought by the physician at that point in the task – in essence, to identify information cues and their purpose during the task. They can also be leveraged to elicit additional knowledge from other SMEs. Miller et al. (2006) described a critiquing process for performing a cognitive task analysis (CTA) that relies on video- and audio-recorded data of a novice performing a task. The results are used to create a 'script' of the novice's performance that can then be critiqued by SMEs. They recorded a novice completing a complex (military intelligence analysis) task, during which the novice was asked to think aloud. Six expert intelligence analysts were read a transcript of the novice's verbalisations while being shown additional material (e.g. screen shots captured, documents accessed and handwritten notes generated by the novice during the task). Experts were asked to comment on the novices' performance as the script was presented. Audio and video recordings, along with handwritten notes of the critiquing process, were used to generate a protocol, which was then analysed to provide insight into how experts approach this task.

However, there are also disadvantages to the use of video. There are some situations, for legal or practical reasons, where the use of video cameras is not possible. For instance, Seagull and Sanderson (2001) were unable to secure permission to videotape surgeries. For studies involving large areas, a single (or even multiple) fixed cameras may be impractical. Setting up, capturing,

storing and converting video data between the variety of available digital and analogue formats (if desired) alone can be costly and time consuming, without even considering the time required for analysis. Sanderson and Fisher (1994) suggest that the time for analysis of sequential observational data, such as that stemming from recorded observations, can reach levels as high as 100 to 1, depending on factors such as the granularity of analysis being conducted (i.e. the length or frequency of activities being recorded, such as gestures or motions vs. movement across a room, or topics raised at a meeting), the specificity of research questions being asked and the complexity of inferences required from the data. Camera angles and distances from the subject can lead to distorted or obstructed images, making it difficult to identify all activities, or detailed interactions with workplace artefacts, reliably. For instance, it may be possible to determine that a person referred to a piece of paper, but not identify what the paper or its content was. Events which occur out of camera coverage cannot be analysed.

These latter cases, in particular, could be alleviated with the supplement of in-person observation or retrospective verbal protocols. While a human observer might not be able to capture all elements of sequential activities in a situation, it is possible for documents to be examined, questions to be asked and activities and equipment to be scrutinized in a more detailed fashion.

## INTERVIEW AND FOCUS GROUP METHODS

Interviews and focus groups are among the most commonly applied knowledge acquisition methods. Unstructured interviews tend to be free form, in which neither the content nor the sequence of the interview topics is predetermined (Cooke, 1994), and are most appropriate early in the knowledge acquisition process, when the analyst is attempting to gain a broad overview of the domain. More typically, analysts will use a semi-structured interview approach in which a list of topics and candidate questions is generated ahead of time, but the specific topics and the order in which they are covered are guided by the responses obtained (e.g. Mumaw et al., 2000; Roth et al., 1999). Structured interview techniques utilize a specific set of questions in a specific order.

A number of semi-structured interview techniques have been developed to uncover the demands of complex cognitive work and the strategies that domain practitioners have developed to meet those demands. A powerful technique is to ground the interviews in analysis of actual past critical incidents so as to understand what made them challenging and why the individuals who confronted the situation succeeded or failed (Flanagan, 1954). The critical decision method (CDM) is a structured interview technique that uses the critical incident technique (Hoffman et al., 1998; Klein and Armstrong, 2005; Klein et al., 1989). CDM focuses on the analysis of challenging situations that the domain practitioner has experienced, and includes four interview phases, or 'sweeps', that examine a past incident in successively greater detail. These phases provide a description of the subtle cues, knowledge, goals, expectancies and expert strategies that domain experts use to handle cognitively challenging situations. CDM has been successfully employed to analyse the basis of experts in a variety of domains, such as fire fighting, neonatal caregiving, and intelligence analysis (Baxter et al., 2005; Hutchins et al., 2003; Klein, 1998).

Concept mapping is another interview technique that is widely used to uncover and document the knowledge and strategies that underlie expertise (Crandall et al., 2006). In concept mapping, the analyst helps domain practitioners build up a representation of their domain knowledge using concept maps, which are directed graphs made up of concept nodes connected by labelled links. They are used to capture the content and structure of domain knowledge that experts employ in problem-solving and decision-making. Concept mapping is typically conducted in group sessions that include multiple domain practitioners (e.g. three to five) and two facilitators. One facilitator provides support in the form of suggestions and probe questions, and the second facilitator creates the concept map based on the participants' comments for all to review and modify. The output is a graphic representation of expert domain knowledge that can be used as input to design of training or decision aids.



Focus groups are interviews conducted with multiple participants. Some special considerations with respect to focus groups include both the number and choice of participants. General recommendations are to limit the size of focus groups to 5–8 participants (Krueger and Casey, 2009), so that everyone has the opportunity to contribute. Some care should be taken, particularly in work settings where participants are likely to know and work with one another, to avoid groups which mix participants of differing status (e.g. employees and managers, physicians and technicians), so that all participants may feel free to contribute openly. Focus groups are best suited for unstructured and semi-structured questioning, which allows participants to add to other participants' comments.

## GENERATIVE METHODS

Generative methods are analysis techniques that invite participants to explore and contribute insight about their lives and work practices to the design process. These methods can be used to better understand user needs and the context in which a product might be used (Stappers et al., 2003). Findings can contribute to the development of new product concepts and the generation of user scenarios and design requirements. With generative methods, participants create, or generate, new material, or artefacts (Sanders, 2000) in response to specific questions posed by researchers. These artefacts can take many forms, including collages, where participants create collections of images and/or words in response to research questions; diaries, where participants independently record information about their activities and practices or Velcro models, where participants create a design for a product or system using elements provided by the researcher (Chung, 2004; Stappers and Sanders, 2003, 2005).

Generative methods are usually guided by an open-ended question motivated by the research topic. For example, Stappers et al. (2003) describe a generative study where participants created a collage showing their experience of home in the past, present and future, and Watts-Perotti et al. (2011) asked mobile workers to create collages showing their current work environment and their ideal work environment. Watts-Perotti et al. comment that

The collage activity allowed participants to examine and describe their work environment in nonverbal ways, and led to a richer description of what they liked about their current environment, and what they would like to change. This part of the study provided details about how work and personal life intermingled, and how participants struggled to maintain a balance in their lives.

After creating the collages, participants described the elements of the collage and explained what these elements meant to them. This discussion was richer than a typical interview, because the collages helped participants become more aware of their work environments, and how they felt about them.

Participants can use personal materials, or those provided by researchers, to create generative artefacts. For example, they could create collages using personal images and objects from home, or using words, images and other materials provided by the researchers. In the case of Velcro modelling, participants typically use material provided by the researchers. Chung (2004) describes a Velcro modelling study that explored ideas for the controls and interface design for microwave ovens. In this study, participants used felt shapes provided by researchers to create an arrangement of controls and displays for a microwave. Velcro modelling is especially good for helping participants create ideas in three-dimensional form (Stappers and Sanders, 2005). For example, Sanders (1993) describes a study where Velcro modelling was used to understand user needs related to the layout of controls for the cab of a military loader vehicle.

Once participants have created the generative materials, they usually verbally present, or explain, the materials to the researchers (Stappers and Sanders, 2005). In this process, participants access and describe parts of their work or experiences that would not typically arise from a natural conversation. Also, as with the Watts-Perotti et al.'s study, the discussion with researchers is grounded

in participants' experiences, because they have spent some time exploring their experience through the creation of the generative materials. It is through this explanation that researchers can better understand participants' personal meaning behind the elements of their materials (i.e. diary entries, collage or mind map elements, etc.).

Analysis of the materials can include qualitative analysis of the explanation of the materials, together with qualitative or quantitative analysis of the created materials. For example, researchers can trace visual or conceptual themes conveyed across materials created by different participants, and/or they can perform statistical analysis on the content of the materials like words or specific images or themes (Stappers and Sanders, 2003).

Generative analysis methods have specific advantages, and are often used to complement other methods of cognitive inquiry. These methods, similar to psychological projective techniques, 'are based on the belief that all people can project and express their needs, wants and aspirations through the use and interpretation of ambiguous visual stimuli' (Stappers and Sanders, 2005). They have the potential to reveal latent unmet needs and aspirations (Chung, 2004), and can provide a direct language for designers, who can in turn incorporate the visual characteristics of the materials into design concepts (Stappers and Sanders, 2005). See Watts-Perotti et al. (2011) for more details about a study that combined generative techniques (diaries and collages) with more traditional cognitive analysis techniques such as observations and interviews to understand the work practices of mobile workers.

## **THEORETICAL PERSPECTIVES AND ANALYSIS FRAMEWORKS**

### **COGNITIVE PSYCHOLOGY AND COGNITIVE SCIENCE**

One prominent approach to cognitive analysis draws on laboratory study methods from the cognitive science and cognitive psychology literature (Ericsson and Simon, 1980; Hoffman, 1987, 2008). This approach has contributed a variety of specific cognitive analysis methods for eliciting and representing the knowledge and strategies that underlie performance at different levels of expertise. This includes the use of 'think-aloud' protocols and the application of psychological scaling methods such as multidimensional scaling of pairwise ratings that provide a means to uncover how experts organize their knowledge (Cooke, 1994; Ericsson and Simon, 1993). It also led to methods for modelling the mental operations that underlie cognitive performance (Gray, 2007).

### **COGNITIVE SYSTEMS ENGINEERING**

Another approach to cognitive analysis has its roots in the cognitive systems engineering (CSE) tradition (Hollnagel and Woods, 1983; Rasmussen, 1986). Analyses of complex accidents, such as the Three Mile Island nuclear power plant incident, revealed that crisis events often involve a confluence of events that have not been previously anticipated. As a consequence, the training, procedures and support systems in place often prove to be inadequate (Sanderson, 2003). One of the major innovations of cognitive systems engineering was the development of functional analysis methods that define the goals, constraints and affordances in a domain that constitute the cognitive problem-space that practitioners need to cope with. The results of the functional analyses are used to develop systems that enable domain practitioners to directly 'perceive' and reason about system goals, constraints and affordances – so as to be able to perform effectively under unanticipated conditions (Hollnagel and Woods, 2005; Rasmussen, 1986; Roth and Woods, 1988; Woods and Hollnagel, 1987, 2006; Woods and Roth, 1988).

Cognitive work analysis (CWA) is the most fully developed analysis framework that grew out of this tradition (Rasmussen, 1986; Vicente, 1999). CWA includes five interlinked analyses that focus successively on different layers of constraints ranging from characteristics of the work domain to cognitive competencies of the individuals engaged in the work. The grounding layer of a CWA is

*a work domain analysis* (WDA) that uncovers and represents the goals, means and constraints in a domain that define the boundaries within which people must reason and act. This provides the basis for identifying functions to be performed by humans (or machines) and the cognitive activities those entail. The remaining layers of the CWA build on the WDA foundation. The *control task analysis* examines what work situations can arise in the domain and defines the functions required to achieve domain goals in those work situations – without specifying what agent (human or machine) will perform the functions. *Strategies analysis* identifies alternative strategies that can be used to perform the control tasks. *Social, organizational and cooperation analyses* explore how work can be distributed across multiple agents (people and/or machines) and organizational structures and the implication for communication and coordination requirements. Finally, *worker competencies analysis* examines the knowledge and skills that are required to efficiently and effectively perform the cognitive and collaborative functions identified through the prior analyses.

There are a number of variants of work domain analysis that have emerged. The applied cognitive work analysis (ACWA) is a comprehensive design methodology that incorporates a domain analysis (Elm et al., 2003, 2008). The output is represented as a functional abstraction network (FAN) which specifies domain goals along with associated processes and system components that support those goals. The FAN is linked to, and provides the basis for, additional stages of analysis including information requirements and representation design. Potter et al. (2003) describe an application of this methodology to military command-and-control, where the FAN was used to represent abstract concepts such as ‘combat power’ as well as high level goals of complying with military law, and socio-political constraints. Subsequent stages of analysis supported the development of innovative displays that visually represented levels of combat power to support commander decision-making.

The work-centred design methodology (Eggleston, 2003; Evenson et al., 2008; Scott et al., 2005) also draws from this tradition. The phrase ‘work-centred’ is intended to highlight that the focus of the analysis is on the demands and broader context of the work. Work-centred design specifically incorporates an analysis of domain goals, characteristics and constraints that impact decision-making as part of the design process.

## NATURALISTIC DECISION-MAKING

Another distinct strand of cognitive analysis addresses decision-making in naturalistic contexts (e.g. Klein, 1998; Klein et al., 1989, 1993). Researchers from this tradition studied expert decision-making in real-world settings (e.g. nurses; fire fighters; military commanders). Their research revealed that expert decision-making in naturalistic settings diverged significantly from the pre-prescriptive decision-making models derived from decision theory, and the choice behaviour observed in artificial laboratory settings.

This line of research made a forceful case for studying decision-making in actual real-world settings and led to development of a variety of new CTA methods for uncovering the knowledge and strategies that underlie practitioner performance (Klein et al., 1989). CTA methods drawing from this tradition include structured interview techniques such as the CDA (Hoffman et al., 1998; Klein et al., 1989), the applied cognitive task analysis method (Militello and Hutton, 1998), the goal-directed task analysis method (Endsley et al., 2003), and cognitive field observation methods that examine performance in actual environments or high fidelity simulators (Roth and Patterson, 2005; Woods, 1993; Chapter 5 of Woods and Hollnagel, 2006).

## ETHNOGRAPHY

Ethnography is a methodological framework that originated in the field of cultural anthropology. Cultural anthropologists use field-based methods like observations and in-context interviews to derive understandings of culture from the insider’s perspective (Hoey, n.d.). Since their inception, ethnographic methods have been adopted by many different fields, including sociology, education

and cognitive engineering, and have come to represent a collection of qualitative methods conducted in the context in which the practices of interest naturally occur. Central to these methods is an emphasis placed on direct observations of daily life and work, to understand the natural organization of activities, environments and interactions, and to explain these elements from the practitioner's point of view (Szymanski and Whalen, 2011). Ethnography is a departure from traditional human factors studies, conducted in laboratories, with an emphasis on controlling variables and explaining the causes of behaviour. Ethnography is fundamentally a descriptive methodology, rather than prescriptive (Blomberg et al., 1993, 2003), though design implications can certainly be derived from the descriptions that are produced. Proponents of ethnography argue that these methods can identify cognitive cues and patterns that might never be considered for inclusion in lab-based studies.

Traditionally, ethnographers in cultural anthropology have spent significant amounts of time with participants, sometimes living with them for months or even years. The budget and time required for this fully immersive approach are not practical for many researchers. In these cases, ethnographers often combine several techniques to get a more comprehensive understanding of participants' lives and practices in a shorter amount of time.

The main foundational methods of traditional ethnography are observations and in-context interviews. Margaret Mead once said that 'what people say, what they do, and what they say they do are entirely different things' (Lewis, n.d.). For this reason, ethnographers collect and compare multiple kinds of data to get a deeper understanding of participants' lives and work practices from the participants' point of view. Often, initial interviews about work practices can lead to a formalized, or ideal, description of how work gets done. Participants may refer to formal diagrams summarizing idealized steps of a process, or organization charts showing the formal reporting structure within their company (Pew et al., 2007). Observations can reveal the hidden work not represented in these descriptions. For example, observations show how participants handle contingencies, and reveal the informal networks that arise to support complex tasks and processes. The critical decision-making methods mentioned earlier can also expand interviews to include discussions about the exceptions to the normalized descriptions of work practices. Often, the contrast between the idealized descriptions and the observed work practices can reveal latent needs and opportunities for new tools or processes to improve performance and user experience.

Ethnographers tend to value the creation of thick descriptions through their inquiry and summary of findings (Geertz, 1973). Thick descriptions explain human behaviour and the context in which it occurs, in ways that make the behaviour meaningful to an outsider. Ethnographers tend to explore beyond the boundaries of narrow research questions, focusing on the complexities of work and life, rather than attempting to boil the findings down into simple cause-and-effect descriptions. Ethnographers consider how the environment affects work practices, how practices change in different contexts and how the work is coordinated across networks of practitioners.

Since ethnographers focus on the complexities of work practices, the details from the data are very important to the analysis of the study. Many ethnographers record their interviews and observations with audio and/or video recorders, so they can review these details during analysis (Wall and Brun-Cottan, 1995). Recording can be especially beneficial in cases where the work is fast paced and complex, and it allows researchers to take fewer notes and focus more attention on their observations or interviews.

## INTERACTION DESIGN

One of the philosophies behind human-computer interaction (HCI) methods such as interaction design is to involve practitioners, or end-users in the design of the systems they will be using. Chapter 13 of this volume covers methods for evaluation of HCI in more depth, but a brief overview of elements of interaction design relevant to cognitive work is presented here. Interaction design focuses on the creation of human-machine system and the interfaces between systems and practitioners, who will use them to accomplish tasks and goals (Evenson et al., 2008). Ideally, these

methods are iterative, engaging users throughout the design process from the early exploratory phase to the evaluative phase where developed designs are refined (Evenson et al., 2008). Involving users throughout the design process grounds the design in the context of use, and increases the likelihood that the final product will be useful and usable for the end-users. See Chapters 11 and 34 for more detailed discussions of inclusive design and participatory ergonomics.

User-centred interaction design encompasses all parts of the design process and includes a wide variety of techniques. For example, in the early stages of design, Blomberg et al. (1993) conducted participatory analysis of exploratory ethnographic studies by asking study participants to review representations and diagrams of observed work practices, which were created by the researchers. The input from participants helped refine and expand the study findings to include a richer understanding of end-users and their context of use. Foster and Gibbons (2007) describe an evaluative study in the later stages of design, where participants helped refine a library home page website. Participants marked up printed versions of the current home page to indicate what they liked, what they would get rid of and what they would add to the home page.

A typical characteristic of user-centred design is that end-users contribute directly to evolving design concepts (Pew and Mavor, 2007). Many researchers accomplish this direct input by asking end-users to contribute their ideas to incomplete prototypes (Pew and Mavor, 2007), or to create representations of their requirements and aspirations using ambiguous stimuli like those described in the generative research section earlier. Generative materials created by participants can provide direct visual input into the design process, in addition to rich descriptions of the context of use (Stappers and Sanders, 2003).

Incomplete prototypes can take an infinite number of forms. These prototypes are often low-fidelity (i.e. not fully functional), and can take the form of physical objects that study participants can manipulate. For example, Wall and Mosher (1994) describe a technique called ‘visualization and construction games’, which uses a static cling game board and vinyl game pieces. The pieces can represent components of technology or of processes (including people, activities and/or technology). ‘The goal is to co-develop an arrangement of the game pieces that represents, for example, an accurate description of a work activity or an agreed upon proposal for a technology concept (Wall and Mosher, 1994)’.

### **Scenario-Based Design**

Scenario-based design is another methodological framework with the goal of embedding user experience into the design process (Carroll, 1995, 2000). The intention of this framework is to expand developers’ attention beyond just functionality development, to include an understanding of how practitioners might use the system being developed (Pew and Mavor, 2007). With this approach, ‘narrative descriptions of envisioned usage scenarios are employed throughout the design process’ (Rosson and Carroll, 2009).

Materials used in scenario-based design can include narratives, storyboards, use cases and stories (Pew and Mavor, 2007). The materials provide descriptions of who the users are, the ways in which systems will be used and the environments in which they will be used. Essentially, they are stories demonstrating what it will be like to use the system being developed. End-users are often involved in the construction and refinement of these scenarios. These materials become reference materials for system designers, and can evolve into use cases, which ‘directly inform design and development on an event-by-event (or action-by-action) basis’ (Pew and Mavor, 2007).

### **User Needs/Requirements Analysis**

Since HCI methods are conducted with the goal of informing user-centred design, the output of these methods often leads to summaries of user needs and requirements for the system being developed. This phase of a project distils findings from exploratory user experience research into actionable information about the constraints that must be considered as a system evolves. Exploratory studies lead to ‘descriptions of the setting, participants, and actions that the system needs to support’



(Evenson et al., 2008). These summaries are synthesized into frameworks or models which summarize elements of user experience which must be accounted for in the design. 'Designers use the models to think with, generate design principles, produce concepts for what might be, and then iteratively refine the resources until a final interface direction is developed, detailed, and released' (Evenson et al., 2008).

## OUTPUTS AND APPLICATIONS OF THE METHODS

Cognitive analyses result in a wide range of outputs used to support the design, implementation and continued refinement of systems, technologies and structures which support complex cognitive work. Cognitive analyses are also used through the design cycle of both existing and first-of-a-kind systems. Other applications include the design of human-centred automation systems, corporate knowledge capture and to generate intellectual property.

## INTERFACES FOR MONITORING AND CONTROL

Guidance for the design of information displays is one of the most common outputs from a cognitive analysis, primarily in the form of recommendations for information requirements, display content and organizational schemes. This guidance is derived from an understanding of, for instance, the information requirements associated with control or monitoring of the various goals, constraints, processes and functions identified through a work domain analysis; critical decision cues identified through cognitive task analyses; or relationships among tasks and system information indicated by expert strategies and task coordination requirements. Examples of successful design applications that have relied on cognitive analysis include redesign of an airborne weapons director station (Klinger and Gomes, 1993); design of next-generation navy ships (Bisantz et al., 2003; Burns et al., 2004); design of next-generation power plants (Burns et al., 2008; Roth et al., 2001); design of auditory displays for anaesthesia monitoring (Watson and Sanderson, 2007) and design of integrated visualizations for command and control airlift mission monitoring and planning (Roth et al., 2006). For example, Bisantz et al. (2003) made recommendations regarding command-and-control information displays based on a CSE analysis of a navy ship. Because the analysis was performed early in the design of the ship, the recommendations consisted of information elements and the functional organization of those elements into display panels, rather than specific interface designs.

Ecological interface design (EID) is a specific interface design method which is based on cognitive systems engineering analyses (Burns and Hajdukiewicz, 2004; Vicente, 2002; Vicente and Rasmussen, 1992). Goals of the method are to design information displays and controls such that users are able to exhibit skilled control or execute procedures during routine operations while having information available to support more effortful cognitive activities, such as fault diagnosis or problem-solving during novel or non-routine circumstances. EID takes as input information gathered during the work domain analysis component of a CWA, and has been used to design interfaces in a wide array of domains such as health monitoring, military command control and process control; see Burns and Hajdukiewicz (2004) for a methodological description as well as extensive examples. Often, EIDs make use of object or configural graphics, where patterns or shapes 'emerge' from the combination of object dimensions. The objects are designed such that object dimensions represent process parameters and the emergent features (such as size, alignment or symmetry) are mapped to key variables.

Researchers of military systems (Bennett et al., 2008; Hall et al., 2012) used the EID methodology to develop a command-and-control display to support army tactical operations. The displays contained information derived from a WDA including representations of important resources (e.g. status and capabilities of friendly units) and abstract relationships such as the force ratio, or relative amount of combat power between friendly and enemy forces. The displays also contained information to support activities such as monitoring the status of the situation relative to a plan, or planning alternative

courses of action. A subsequent comparison of the EID interface to a more conventional interface during laboratory-based military scenarios showed that the EID was more effective across a number of measures including accuracy of information reported about the ongoing situations (Hall et al., 2012).

Jamieson and colleagues used both work domain and task analyses to create an ecological interface for a complex petrochemical refining process (Jamieson, 2007; Jamieson et al., 2007). The display included graphical representations of flows and their relationships, as well as configurable overview graphics showing variables that are critical to monitor. An evaluation of these displays compared to standard petrochemical displays showed advantages for the EID displays on measures such as number of control actions and improved fault diagnoses.

## **SUPPORT THROUGHOUT THE DESIGN CYCLE**

Cognitive analysis has been successfully used to impact the design, implementation and refinement of complex human-technology systems throughout the design cycle. For instance, work domain analyses have been employed to provide early input into the human systems integration (HSI) issues in a number of large-scale, first-of-a-kind projects including the design of a next-generation power plant (Roth et al., 2001); the design of a next-generation U.S. Navy Battleship (Bisantz et al., 2003; Burns et al., 2004) and the design of a next generation Canadian Frigate (Burns et al., 2000). Naikar and her colleagues have been applying CWA methods to the design of a first-of-a-kind military aircraft (Naikar, 2009; Naikar and Sanderson, 1999, 2001; Naikar and Saunders, 2003; Naikar et al., 2003; Sanderson, 2003; Sanderson et al., 1999). Their work has impacted system design throughout the design cycle, including the use of cognitive engineering methods to evaluate alternative design proposals offered by different vendors; determining the most appropriate composition of crew members and functions to staff the aircraft; the definition of needs for training and training simulators and the assessment of risks associated with upgrading existing defence platforms.

There are also numerous examples of cognitive engineering analysis being used to support the design of social and organizational aspects of work. For instance, Jenkins et al. (2008a) applied work domain analysis techniques to understand how helicopter mission planning work could be distributed across multiple team members. Chin et al. (1999) applied similar techniques to illustrate the scope of command and control and reporting tasks across different levels in a military hierarchy. Gualitieri et al. (2000) applied work domain analysis methods to illustrate span of responsibility across nuclear power plant crew members and how it shifts to accommodate shifting demands across a variety of normal and emergency plant conditions.

CTA methods have also been applied at and beyond the point when systems are fielded. Implementation of new technology necessarily impacts practitioner activities and leads to the development of new tasks and contexts for work (Carroll and Campbell, 1989). Often, these changes cannot be anticipated prior to implementation (Patterson et al., 2002; Woods and Dekker, 2000). For example, O'Hara and Roth (2005) discovered that newly implemented computerized procedures at a power plant caused (as expected) primary operators to have better situation awareness of plant state while unexpectedly reducing awareness of others who now had less reason to be kept in-the-loop. New training methods were subsequently developed to capitalize on the crew members' freed-up mental resources (they could now provide an independent and diverse check on plant state), resulting in improved shared situation awareness of the entire team.

## **HUMAN-CENTRED AUTOMATION**

Cognitive analysis methods have also been employed to inform the design of human-centred automation. For example, Roth and Multer (2009) used cognitive field observation and focus group interview methods to explore the impact of new forms of train control automation on the cognitive performance of train crews. Macbeth et al. (2012) used a hybrid CTA method that incorporates decision ladders to identify decision-making tasks that must be augmented with automation for design

of support for an unmanned vehicle supervisory control task. Adams et al. (2009) used multiple CTA methods including field observations, a goal-directed task analysis and elements of a CWA to guide the design of unmanned aerial vehicles as well as to define the roles and responsibilities associated with effectively integrating the technology into existing wilderness search and rescue operations. Scott and colleagues (DePass et al., 2011; Truxler et al., 2012) have also employed CTA methods as part of a work-centred approach to design of collaborative automation in support of military transportation planning and scheduling.

### **CORPORATE KNOWLEDGE CAPTURE**

Another application area for CTA is corporate knowledge capture (Hoffman, 2008). There has been growing concern in many organizations of losing critical corporate 'know-how' as experienced personnel retire or move on to other positions (Hoffman and Hanes, 2003). CTA methods can be used to capture this critical knowledge for the purposes of preserving and sharing it with younger, less experienced personnel. Hoffman and Moon (2010) report several successful projects where CTA methods, that have included concept maps and the CDA, have been used to elicit expert domain knowledge and reasoning strategies of retiring experts at utilities and power plants. These have included eliciting and modelling domain expertise as varied as knowledge and strategies for reactor design, steam turbine repair and business relations.

### **INTELLECTUAL PROPERTY**

The methods used to study and analyse complex cognitive work have great potential for creating intellectual property. Often, the main goal of these methods is to identify inefficiencies and latent user needs in current systems, in order to generate new solutions to solve these problems. Because they are inspired by actual practice, and are aimed at solving problems that have not yet been solved, these new solutions have the potential to be unique, original and relevant to practitioners. These same methods also provide rich descriptions of the context of use, which can explain why the intellectual property is relevant, and help define the detailed features of an invention, and the scenarios demonstrating how the invention could be used.

One example of a patent that was informed by ethnographic research is U.S. patent number 7,894,111, titled 'Method and system to determine preferred colour management settings for printing documents' (Woolfe et al., 2011). Findings from the ethnographic research revealed that practitioners were not using digital production printing presses in the ways in which the designers expected (O'Neill et al., 2007). The designers of the system created the presses to work with standardized digital colour profiles, which were supposed to be attached to the files submitted for printing. However, the practitioners who created these files were not attaching the expected colour profiles; therefore, the press operators were left with the task of manually adjusting the colour settings on the press. Since the colour controls on the press were designed to work with colour profiles, which are mathematically complex, these adjustments were done mostly through trial and error, which required a significant time investment (Sprague et al., 2007).

Based on the findings of this study, Woolfe et al. (2011) developed an invention proposing a new tool to support the quick adjustment of the colour controls on the press. In this invention, practitioners are presented with a range of colour options, printed by the actual press on one page, so practitioners can base their colour adjustments on a comparison of the options. The practitioners were already using this method as a workaround when colour profiles were not available, but the presses did not provide an easy way to print the range of colour options. Instead, practitioners created multiple prints through trial and error based on a set of heuristics for how to approximate the range they were trying to explore. This invention allows practitioners to shorten this comparison task by automatically generating the range, using algorithms that take advantage of the current design of the colour controls on the press (Woolfe et al., 2011).



It was the detailed study of the inefficiencies of the current process, and the difficulties experienced by actual practitioners, which led to the creation of this invention. The steps of the invention were inspired by an understanding of why the system was not working, and also by how practitioners created workarounds to compensate for the failures of the design assumptions. Study findings also contributed to the description of why the invention was relevant (i.e. the background description), how it was unique from similar inventions and how it would be used.

## TECHNIQUES FOR MAKING AN IMPACT

There are many ways to make an impact with CWA studies. If the work is exploratory, the study results may contribute to brainstorming sessions to develop new products, or may affect business or development strategy. The study by Watts-Perotti et al. (2011) provided user requirements for Xerox's mobile print product line and contributed to corporate strategy by revealing how paper is currently used as a compensation for technology weakness (Watts-Perotti et al., 2011).

Often, cognitive analyses are conducted with the goal of contributing to the development or refinement of a specific system, rather than being exploratory like the Future of Work study (Watts-Perotti et al., 2011). In these cases, researchers may be part of the design and development team, or they may be loosely coupled with other stakeholders like developers, designers and marketers. Either way, the information must be translated from study results into system design and development. In the loosely coupled scenarios, researchers may spend more time coordinating buy-in from other stakeholders.

Two ways to increase the usefulness and relevance of the methods discussed here are to create multidisciplinary research teams, with invitations for stakeholders to participate, and to create engaging representations of study findings, which resonate with stakeholders (Wall and Mosher, 1994).

## MULTIDISCIPLINARY TEAMS

Multidisciplinary teams increase the robustness of the materials that are produced (Watts-Perotti and Woods, 2007). Evenson et al. (2008) describe a team composed of cognitive engineers, interface design specialists and software engineers. They note that 'the interplay among these perspectives was critical in identifying opportunities for support that were simultaneously grounded in an understanding of what is needed and what is possible and translating those support requirements into effective visualizations' (Evenson et al., 2008). O'Neill et al. (2011) note that when ethnographers worked closely with designers and developers in their project, the ongoing design was 'consistently grounded in the field study findings'. Multidisciplinary teams are more robust, because a broader set of constraints can be integrated into the development process. In Evenson et al.'s case, 'the fact that the work was performed within the larger systems engineering context of an integrated product team enabled larger systems concerns and stakeholder perspectives to be identified and addressed' (Evenson et al., 2008).

When it is not possible to create an integrated multidisciplinary team, researchers can still recover some multidisciplinary advantages by inviting stakeholders to participate in the research. Watts-Perotti et al. (2011) created an advisory board to facilitate their interactions with business groups across the company. The board included stakeholders from key business groups, who could potentially make use of study findings. Board members provided input into all stages of the study, contributing questions that influenced the study focus and analysis, accompanying researchers into the field and participating in concept brainstorming and evaluation (Watts-Perotti et al., 2011). They note that

In addition to increasing visibility of the project, the advisory board also helped us create resonance for the project within the company. In gaining exposure to questions and concerns from various stakeholders, we were able to not only answer their questions, but push them further than they might have expected, by challenging some of the assumptions that resided behind their questions.

Another way to involve stakeholders in cognitive analyses is to invite SMEs into the field. In Colombino et al.'s study (2011), the team invited an imaging scientist into the field to determine how digital printing press operators were preparing the press to print digital files. Because they knew the rationale behind the design of the press colour controls, the imaging scientist helped the team discover that the system was not being used as designers had expected. [This discovery was very surprising to him, and he became a translator and advocate for the research findings during technical meetings discussing future design refinements. He also collaborated with the team to develop the invention mentioned earlier (Woolfe et al., 2011).]

## REPRESENTING STUDY FINDINGS AND IMPLICATIONS

Another way of making an impact with complex work analysis methods is to go beyond descriptive summaries to create engaging representations of study findings which resonate with stakeholders. Wall and Mosher (1994) provide an excellent overview of tools and techniques, including a work practice toolkit, which is used in Xerox to create concrete representations of observed workflows, to facilitate conversations with designers and developers (see Martin and Wall, 2011; Sprague et al., 2011 for examples). Watts-Perotti et al. (2011) represented their findings in a design directions document, which contained high-level concepts and scenarios that were mapped to study findings. 'The concepts were not specific recommendations for product ideas, but were instead high level templates that intentionally left room for readers to fill in details that resonated with their business function' (Watts-Perotti et al., 2011).

Video is another way to communicate study findings in ways that engage and resonate with stakeholders. Szymanski et al. (2011) created a 'Day in the life of a call centre agent' documentary video to demonstrate the complexity of call centre work. Researchers at Palo Alto Research Center created a video podcast summarizing findings from a study of digital projector usage (PARCinc., 2009). Watts-Perotti et al. (2011) also created a video podcast to summarize findings from their study to increase the visibility of their project so that viewers would remember results more easily.

## CONCLUSIONS

There is a great variety in the theoretical perspectives, data collection methods and analytic frameworks which are useful in the analysis of complex, cognitive work. These perspectives are linked by common goals of understanding work domain complexities as well as knowledge and strategies of experienced domain practitioners. This chapter provides an overview of the rich diversity of cognitive analyses techniques, as well as the wide range of applications for the methods.

## ACKNOWLEDGEMENTS

This chapter includes material drawn from a previous edition of Bisantz and Drury (2005); as well as material adapted from Bisantz and Roth (2008).

## REFERENCES

- Adams, J. A., Humphrey, C. M., Goodrich, M. A., Cooper, J. L., Morse, B. S., Engh, C. and Rasmussen, N. (2009). Cognitive task analysis for developing unmanned aerial vehicle wilderness search support. *Journal of Cognitive Engineering and Decision Making*, 3(1), 1–26.
- Annett, J. (2000). Theoretical and pragmatic influences on task analysis methods. In Schraagen, J. M., Chipman, S. F. and Shalin, V. L. (Eds.), *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Annett, J. and Duncan, K. D. (1967). Task analysis and training design. *Occupational Psychology*, 41, 211–221.
- Artman, H. (2000). Team situation assessment and information distribution. *Ergonomics*, 8, 1111–1128.
- Bainbridge, L. and Sanderson, P. (1995). Verbal protocol analysis. In J. R. Wilson and E. N. Corlett (Eds.), *Evaluation of Human Work* (pp. 169–201). London, U.K.: Taylor & Francis Group.

- Baxter, G. D., Monk, A. F., Tan, K., Dear, P. R. F. and Newell, S. J. (2005). Using cognitive task analysis to facilitate the integration of decision support systems into the neonatal intensive care unit. *Artificial Intelligence in Medicine*, 35, 243–257.
- Bennett, K. B., Posey, S. M. and Shattuck, L. G. (2008). Ecological interface design for military command and control. *JCEDM*, 2(4), 349–385.
- Beyer, H. and Holtzblatt, K. (1998). *Contextual Design: Defining Customer-Centered Systems*. New York: Morgan Kaufmann Publishers.
- Bisantz, A. M. and Drury, C. G. (2005). Applications of archival and observational data. In J. R. Wilson and N. Corlett (Eds.), *Evaluation of Human Work* (3rd ed., pp. 61–82). Boca Raton, FL: Taylor & Francis Group.
- Bisantz, A. M., Pennathur, P., Guarrera, T. K., Fairbanks, R. J., Perry, S. J., Zwemer, F. and Wears, R. L. (2010). Emergency department status boards: A case study in information systems transition. *Journal of Cognitive Engineering and Decision Making*, 4(1), 39–68.
- Bisantz, A. M. and Roth, E. M. (2008). Analysis of cognitive work. In D. A. Boehm-Davis (Ed.), *Reviews of Human Factors and Ergonomics*, 3, 1–43. Santa Monica, CA: Human Factors and Ergonomics Society.
- Bisantz, A. M., Roth, E. M., Brickman, B., Gosbee, L., Hettinger, L. and McKinney, J. (2003). Integrating cognitive analyses in a large scale system design process. *International Journal of Human Computer Studies*, 58, 177–206.
- Blomberg, J., Burrell, M. and Guest, G. (2003). An ethnographic approach to design. In J. Jacko and A. Sears (Eds.), *Handbook of HCI*. Mahwah, NJ: Erlbaum.
- Blomberg, J., Giacomi, J., Mosher, A. and Swenton-Wall, P. (1993). Ethnographic field methods and their relation to design. In D. Schuler and A. Namioka (Eds.), *Participatory Design: Principles and Practice*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bowers, C. A., Jentsch, F., Salas, E. and Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors*, 40(4), 672–679.
- Burns, C. M., Bisantz, A. M. and Roth, E. M. (2004). Lessons from a comparison of work domain models: Representational choices and their implications. *Human Factors*, 46, 711–727.
- Burns, C. M., Bryant, D. and Chalmers, D. (2000). A work domain model to support naval command and control. In *Proceedings of the 2000 IEEE International Conference on Systems, Man and Cybernetics* (pp. 2228–2233). IEEE, Piscataway, NJ.
- Burns, C. M. and Hajdukiewicz, J. (2004). *Ecological Interface Design*. Boca Raton, FL: CRC Press.
- Burns, C. M., Skraaning Jr., G., Jamieson, G., Lau, N., Kwok, J., Welch, R. and Andresen, G. (2008). Evaluation of ecological interface design for nuclear process control: Situation awareness effects. *Human Factors*, 50(4), 663–679.
- Burns, C. M. and Vicente, K. J. (2000). A participant-observer study of ergonomics in engineering design: How constraints drive design process. *Applied Ergonomics*, 31, 73–82.
- Carroll, J. M. (1995). *Scenario Based Design: Envisioning Work and Technology in System Development*. New York: Wiley & Sons.
- Carroll, J. M. (2000). *Making Use: Scenario Based Design of Human-Computer Interactions*. Cambridge, MA: MIT Press.
- Carroll, J. M. and Campbell, R. L. (1989). Artefacts as psychological theories: The case of human-computer interaction. *Behaviour and Information Technology*, 8(4), 247–256.
- Chin, M., Sanderson, P. and Watson, M. (29 June–1 July 1999). Cognitive work analysis of the command and control work domain. In *Proceedings of the 1999 Command and Control Research and Technology Symposium*. Newport, RI: United States Naval War College.
- Chung, W. (18–20 March 2004). The method and mode of research: Letting people's aspirations drive innovation and team collaboration. In *Education That Works: The NCIAA Eighth Annual Meeting* (pp. 237–240).
- Colombino, T., O'Neill, J., Martin, D., Grasso, A., Wilamowski, J., Roulland, F., Sprague, M.A. and Watts-Perotti, J. (2011). Seeing the right colour: Technical and practical solutions to the problem of accurate colour reproduction in the digital print industry. In P. Szymanski and J. Whalen (Eds.), *Making Work Visible: Ethnographically Grounded Case Studies of Work Practice*. Cambridge, U.K.: Cambridge University Press.
- Cook, R. I. and Woods, D. D. (1996). Adapting to new technology in the operating room. *Human Factors*, 38(4), 593–613.
- Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41, 801–849.
- Crandall, B., Klein, G. A. and Hoffman, R. R. (2006). *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. Cambridge, MA: The MIT Press.

- Degani, A. and Wiener, E. (1993). Cockpit checklists: Concepts, design, and use. *Human Factors*, 35(2), 345–360.
- Degani, A. and Wiener, E. (1997). Procedures in complex systems: The airline cockpit. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 27(3), 302–312.
- DePass, B., Roth, E. M., Scott, R., Wampler, J. L., Truxler, R. and Guin, C. (31 May–3 June 2011). Designing for collaborative automation: A course of action exploration tool for transportation planning. In *Proceedings of the 10th International Conference on Naturalistic Decision Making* (pp. 95–100), Orlando, FL.
- Eggleson, R. G. (2003). Work-centered design: A cognitive engineering approach to system design. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 263–267). Santa Monica, CA: Human Factors and Ergonomics Society.
- Elm, W. C., Gualtieri, J. W., McKenna, B. P., Tittle, J. S., Peffer, J. E. and Szymczak, S. S. (2008). Integrating cognitive systems engineering throughout the systems engineering process. *Journal of Cognitive Engineering and Decision Making*, 2, 249–273.
- Elm, W. C., Potter, S. S., Gualtieri, J. W., Easter, J. R. and Roth, E. M. (2003). Applied cognitive work analysis: A pragmatic methodology for designing revolutionary cognitive affordances. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 357–382). Mahwah, NJ: Lawrence Erlbaum Associates.
- Endsley, M. R., Bolte, B. and Jones, D. G. (2003). *Designing for Situation Awareness: An Approach to User-Centered Design*. CRC Press, Boca Raton, FL.
- Ericsson, K. A. and Simon, H. A. (1980). *Verbal Reports as Data*. Psychological Review, 87(3), 215–251.
- Ericsson, K. A. and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.
- Evenson, S., Muller, M. and Roth, E. M. (2008). Capturing the context of use to inform system design. *Journal of Cognitive Engineering and Decision Making*, 2(3), 181–203.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358.
- Foster, N. F. and Gibbons, S. (2007). *Studying Students: The Undergraduate Research Project at the University of Rochester*. Chicago, IL: Association of College and Research Libraries.
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The Interpretation of Cultures: Selected Essays* (pp. 3–30). New York: Basic Books.
- Gilbreth, F. and Gilbreth, L. (1917). *Applied Motion Study*. London, U.K.: Sturgis & Walton.
- Glaser, B. G. and Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine Publishing Company.
- Gorman, M. E. and Militello, L. G. (2004). Internet searching by ear: Decision flow diagrams for sightless internet users. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 243–247). Santa Monica, CA: Human Factors and Ergonomics Society.
- Gray, W. D. (Ed.). (2007). *Integrated Models of Cognitive Systems*. New York: Oxford University Press.
- Gualtieri, J., Roth, E. and Eggleson, R. (30 April–4 May 2000). Utilizing the abstraction hierarchy for role allocation and team structure design. In *Proceedings of the Fifth International Conference on Human Interaction with Complex Systems*. Urbana, IL: US Army Research Laboratory.
- Hall, D. S., Shattuck, L. and Bennett, K. B. (2012). Evaluation of an ecological interface design for military command and control. *Journal of Cognitive Engineering and Decision Making*, 6(2), 165–193.
- Hammersley, M. and Atkinson, P. (1983). *Ethnography Principles in Practice*. London, U.K.: Tavistock Publications.
- Hoc, J. M. and Leplat, J. (1983). Evaluation of different modalities of verbalization in a sorting task. *International Journal of Man-Machine Studies*, 18, 283–306.
- Hoey, B. (n.d.). What is ethnography? Website: [http://www.brianhoey.com/General%20Site/general\\_defn-ethnography.htm](http://www.brianhoey.com/General%20Site/general_defn-ethnography.htm). <http://brianhoey.com/research/ethnography/>. (Last accessed 24 Nov 2014.)
- Hoffman, R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *AI Magazine*, 8(Summer), 53–67.
- Hoffman, R. (2008). Human factors contributions to knowledge elicitation. *Human Factors*, 50(3), 481–488.
- Hoffman, R., Crandall, B. and Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, 40(2), 254–276.
- Hoffman, R. and Hanes, L. F. (July–August 2003). The boiled frog problem. *IEEE Intelligent Systems*, 18, 68–71.
- Hoffman, R. and Lintern, G. (2006). Eliciting and representing the knowledge of experts. In K. A. Ericsson, N. Charness, P. Feltovich and R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 203–222). New York: Cambridge University Press.
- Hoffman, R. and Militello, L. G. (2008). *Perspectives on Cognitive Task Analysis: Historical Origins and Modern Communities of Practice*. Boca Raton, FL: CRC Press/Taylor & Francis Group.

- Hoffman, R. and Moon, B. (7–11 November 2010). Knowledge capture for the utilities. In *Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies NPIC&HMIT 2010*, Las Vegas, NV, on CD-ROM. LaGrange Park, IL: American Nuclear Society.
- Hollnagel, E. and Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies*, 18, 583–600.
- Hollnagel, E. and Woods, D. D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton, FL: Taylor & Francis Group.
- Hutchins, S. G., Pirolli, P. and Card, S. (2003). Use of critical analysis method to conduct a cognitive task analysis of intelligence analysts. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 478–482). Santa Monica, CA: The Human Factors and Ergonomics Society.
- Jamieson, G. A. (2007). Ecological interface design for petrochemical process control: An empirical assessment. *IEEE Transactions on Systems, Man and Cybernetics*, 37, 906–920.
- Jamieson, G. A., Miller, C. M., Ho, W. H. and Vicente, K. J. (2007). Integrating task and work domain based work analyses in ecological interface design: A process control case study. *IEEE Transactions on Systems, Man and Cybernetics*, 37, 887–905.
- Jenkins, D. P., Stanton, N. A., Salmon, P. M., Walker, G. H. and Young, M. S. (2008a). Using cognitive work analysis to explore activity allocation within military domains. *Ergonomics*, 51, 798–815.
- Kirschenbaum, S. S. (2004). The role of comparison in weather forecasting: Evidence from two hemispheres. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 306–310). Santa Monica, CA: Human Factors and Ergonomics Society.
- Klein, G. A. (1998). *Sources of Power: How People Make Decisions*. Cambridge, MA: The MIT Press.
- Klein, G. A. and Armstrong, A. A. (2005). Critical decision method. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas and H. Hendrick (Eds.), *Handbook of Human Factors and Ergonomics Methods* (pp. 35.1–35.8). Boca Raton, FL: CRC Press.
- Klein, G. A., Calderwood, R. and MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(3), 462–472.
- Klein, G. A., Orasanu, J., Calderwood, R. and Zsombok, C. E. (Eds.). (1993). *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation.
- Klinger, D. W. and Gomes, M. G. (1993). A cognitive systems engineering application for interface design. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 16–20). Santa Monica, CA: Human Factors and Ergonomics Society.
- Krueger, R. A. and Casey, M. A. (2009). *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Pine Forge Press.
- Lewis, J. J. (n.d.). Margaret mead quotes. Website: [http://womenshistory.about.com/cs/quotes/a/qu\\_margaretmead.htm](http://womenshistory.about.com/cs/quotes/a/qu_margaretmead.htm). (Last accessed 24 Nov 2014.)
- Macbeth, J. C., Cummings, M. L., Bertuccelli, L. F. and Surana, A. (2012). Interface design for unmanned vehicle supervision through hybrid cognitive task analysis. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Mackenzie, C. F. and Xiao, Y. (2011). Video analysis: An approach for use in health care. In P. Carayon (Ed.), *Handbook of Human Factors and Ergonomics in Health Care and Patient Safety* (2nd ed., pp. 523–542). Boca Raton, FL: CRC Press.
- Martin, N. and Wall, P. (2011). Behind the scenes: The business side of medical records. In P. Szymanski and J. Whalen (Eds.), *Making Work Visible: Ethnographically Grounded Case Studies of Work Practice*. Cambridge, U.K.: Cambridge University Press.
- Militello, L. G. and Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioners toolkit for understanding task demands. *Ergonomics*, 41(11), 1618–1641.
- Miller, J. E., Patterson, E. S. and Woods, D. D. (2006). Elicitation by critiquing as a cognitive task analysis methodology. *Cognition, Technology and Work*, 8, 90–102.
- Montague, E., Xu, J., Asan, O., Chen, P., Chewning, B. and Barrett, B. (2011). Modeling eye gaze patterns in clinician-patient interaction with lag sequential analysis. *Human Factors*, 53(5), 502–516.
- Mumaw, R. J., Roth, E. M., Vicente, K. J. and Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human Factors*, 42(1), 36–55.
- Naikar, N. (2009). Beyond the design of ecological interfaces: Applications of work domain analysis and control task analysis to the evaluation of design proposals, team design and training. In A. M. Bisantz and C. M. Burns (Eds.), *Applications of Cognitive Work Analysis*. Boca Raton, FL: CRC Press.
- Naikar, N., Pearce, B., Drumm, D. and Sanderson, P. (2003). Designing teams for first-of-a-kind complex systems using the initial phases of cognitive work analysis: A case study. *Human Factors*, 42(2), 202–217.



- Naikar, N. and Sanderson, P. M. (1999). Work domain analysis for training-system definition and acquisition. *International Journal of Aviation Psychology*, 9(3), 271–290.
- Naikar, N. and Sanderson, P. M. (2001). Evaluating design proposals for complex systems with work domain analysis. *Human Factors*, 43(4), 529–542.
- Naikar, N. and Saunders, A. (2003). Crossing the boundaries of safe operation: An approach for training technical skills in error management. *Cognition, Technology and Work*, 5, 171–180.
- Nardi, B. A. (1997). The use of ethnographic methods in design and evaluation. In M. Helander, T. K. Landauer and P. V. Prabhu (Eds.), *Handbook of Human Computer Interaction* (pp. 361–367). Amsterdam, the Netherlands: Elsevier Science-North Holland.
- O'Hara, J. M. and Roth, E. M. (2005). Operational concepts, teamwork, and technology in commercial nuclear power stations. In C. Bowers, E. Salas and F. Jentsch (Eds.), *Creating High-Tech Teams: Practical Guidance on Work Performance and Technology* (pp. 139–159). Washington, DC: American Psychological Association.
- O'Neill, J., Martin, D., Colombino, T., Watts-Perotti, J., Sprague, M. A. and Woolfe, G. (2007). Asymmetrical collaboration in print shop-customer relationships. In L. Bannon, I. Wagner, C. Gutwin, R. Harper and K. Schmidt (Eds.), *ECSCW'07: Proceedings of the 10th European Conference on Computer Supported Cooperative Work*, Limerick, Ireland.
- O'Neill, J., Tolmie, P., Castellani, S., Grasso, A. and Roulland, F. (2011). Ethnographically informed technology for remote help giving. In P. Szymanski and J. Whalen (Eds.), *Making Work Visible: Ethnographically Grounded Case Studies of Work Practice*. Cambridge, U.K.: Cambridge University Press.
- PARCinc. (2009). Ethnographic study of projector use. Retrieved from [http://www.youtube.com/watch?v=\\_Mlk-QSCFnU](http://www.youtube.com/watch?v=_Mlk-QSCFnU). (Last accessed 24 Nov 2014.)
- Patterson, E. S., Cook, R. I. and Render, M. (2002). Improving patient safety by identifying side effects from introducing bar coding in medical administration. *Journal of the American Medical Informatics Association*, 9, 540–553.
- Patterson, E. S. and Woods, D. D. (2001). Shift changes, updates, and the on-call architecture in space shuttle mission control. *Computer Supported Cooperative Work*, 10, 317–346.
- Pew, R. W. and Mavor, A. S. (Eds.). (2007). *Human-System Integration in the System Development Process: A New Look*. Washington, DC: The National Academies Press.
- Potter, S. S., Gualtieri, J. W. and Elm, W. C. (2003). Case studies: Applied cognitive work analysis in the design of innovative decision support. In E. Hollnagel (Ed.) *Handbook of Cognitive Task Design* (pp. 653–678). Mahwah, NJ: Lawrence Erlbaum Associates.
- Potter, S. S., Roth, E. M., Woods, D. and Elm, W. C. (2000). Bootstrapping multiple converging cognitive task analysis techniques for system design. In J. M. Schraagen, S. F. Chipman and V. L. Shalin (Eds.), *Cognitive Task Analysis*. Mahwah, NJ: Erlbaum.
- Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An approach to Cognitive Engineering*. New York: North-Holland.
- Rasmussen, J., Pejtersen, A. M. and Goodstein, L. P. (1994). *Cognitive Systems Engineering*. New York: Wiley and Sons.
- Rosson, M. and Carroll, J. (2009). In A. Sears and J. Jacko (Eds.), *Human-Computer Interaction: Development Process Scenario-based Design* (pp. 145–164). Boca Raton, FL: CRC Press.
- Roth, E. M., Christian, C. K., Gustafson, M., Sheridan, T. B., Dwyer, K., Gandhi, T. K., Zinner, M. J. and Dierks, M. M. (2004). Using field observations as a tool for discovery: Analyzing cognitive and collaborative demands in the operating room. *Cognition, Technology and Work*, 6, 148–157.
- Roth, E. M., Lin, L., Kerch, S., Kenney, S. J. and Sugibayashi, N. (2001). Designing a first-of-a kind group view display for team decision making: A case study. In E. Salas and G. Klein (Eds.), *Linking Expertise and Naturalistic Decision Making* (pp. 113–135). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Roth, E. M., Malsch, N., Multer, J. and Coplen, M. (1999). Understanding how train dispatchers manage and control trains: A cognitive analysis of a distributed planning task. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Roth, E. M. and Multer, J. (2009). Technology implications of a cognitive task analysis for locomotive engineers, DOT/FRA/ORD-09/03. Washington, DC: U.S. Department of Transportation/Federal Railroad Administration.
- Roth, E. M., Multer, J. and Raslear, T. (2006). Shared situation awareness as a contributor to high reliability performance in railroad operations. *Organization Studies*, 27(7), 967–987.
- Roth, E. M. and Patterson, E. S. (2005). Using observational study as a tool for discovery: Uncovering cognitive and collaborative demands and adaptive strategies. In H. Montgomery, R. Lipshitz and B. Brehmer (Eds.), *How Professionals Make Decisions* (pp. 379–393). Mahwah, NJ: Lawrence Erlbaum Associates.

- Roth, E. M., Scott, R., Deutsch, S., Kuper, S., Schmidt, V., Stilson, M. et al. (2006). Evolvable work-centered support systems for command and control: Creating systems users can adapt to meet changing demands. *Ergonomics*, 49(7), 688–705.
- Roth, E. M., Stilson, M., Scott, R., Whitaker, R., Kazmierczak, T., Thomas-Meyers, G. and Wampler, J. (2006). Work-centered design and evaluation of a C2 visualization aid. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 255–259). Santa Monica, CA: Human Factors and Ergonomics Society.
- Roth, E. M. and Woods, D. D. (1988). Aiding human performance I: Cognitive analysis. *Le Travail Humain*, 51, 39–64.
- Sachs, P. (1995). Transforming work: Collaboration, learning, and design. *Communications of the ACM*, 38(9), 37–44.
- Sanders, E. B.-N. (1993). Product development research for the 1990's. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 37(5), 422–426. Santa Monica, CA: Human Factors and Ergonomics Society.
- Sanders, E. B.-N. (2000). Generative tools for co-designing. *Co-Designing*, 2000, 3–12.
- Sanderson, P. M. (2003). Cognitive work analysis. In J. M. Carroll (Ed.), *HCI Models, Theories, and Frameworks: Toward a Multi-Disciplinary Science*. San Francisco, CA: Morgan Kaufmann Publishers.
- Sanderson, P. M. and Fisher, C. (1994). Exploratory sequential data analysis. *Human Computer Interaction*, 9(3), 251–317.
- Sanderson, P., Naikar, N., Lintern, G. and Goss, S. (1999). Use of cognitive work analysis across the system life cycle: Requirements to decommissioning. In *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 318–322). Santa Monica, CA: Human Factors and Ergonomics Society.
- Scott, R., Roth, E. M., Deutsch, S. E., Malchiodi, E., Kazmierczak, T., Eggleston, R., Kuper, S. R. and Whitaker, R. (2005). Work-centered support systems: A human-centered approach to intelligent system design. *IEEE Intelligent Systems*, 20(2), 73–81.
- Seagull, F. J. and Sanderson, P. M. (2001). Anaesthesia alarms in context: An observational study. *Human Factors*, 43(1), 66–78.
- Seagull, F. J. and Xiao, Y. (2001). Using eye-tracking video data to augment knowledge elicitation in cognitive task analysis. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 400–403). Santa Monica, CA: Human Factors and Ergonomics Society.
- Sharples, S., Millen, L., Golightly, D. and Balfe, N. (2011). The impact of automation in rail signalling operations. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 225(2), 179–191.
- Sprague, M. A., Woolfe, G., Watts-Perotti, J., Martin, D., Colombino, T. and O'Neill, J. (2007). Ethnographic Studies of Digital Prepress in Color Workflow. in *Proceedings for the Annual Meeting for the Inter-Society Color Council*, Kansas City, MO.
- Sprague, M. A., Martin, N. and Koomen, H. (2011). New ways of working. Implications of work practice transitions. In P. Szymanski and J. Whalen (Eds.), *Making Work Visible: Ethnographically Grounded Case Studies of Work Practice*. Cambridge, U.K.: Cambridge University Press.
- Stappers, P. J. and Sanders, E. B.-N. (2003). Generative tools for context mapping: Tuning the tools. In D. McDonagh, P. Hekkert, J. van Erp and D. Gyu (Eds.), *Design and Emotion*. Boca Raton, FL: CRC Press.
- Stappers, P. J. and Sanders, E. B. N. (2005). Tools for designers, products for users? The role of creative design techniques in a squeezed-in design process. In F. Hsu (Ed.), *Proceedings of the International Conference on Planning and Design*, NCKU, Taiwan, China.
- Stappers, P. J., Sleeswijk Visser, F. and Keller, A. I. (2003). Mapping the experiential context of product use: Generative techniques beyond questions and observations. In *Proceedings of Asian Design International Conference, Asian Society for the Science of Design*, Tsukuba, Japan.
- Suchman, L. and Trigg, R. H. (1991). Understanding practice: Video as a medium for reflection and design. In J. Greenbaum and M. Kyng (Eds.), *Design at Work: Cooperative Design of Computer Systems* (pp. 65–89). Hillsdale, NJ: Lawrence Erlbaum.
- Szymanski, P., Plurkowski, L., Wall, P. and Englert, J. (2011). How can I help you today? The knowledge work of call center agents, In the *Proceedings of the 12th Participatory Design Conference*, August 12–16, Roskilde University, Denmark.
- Szymanski, P. and Whalen, J. (2011). *Making Work Visible: Ethnographically Grounded Case Studies of Work Practice*. Cambridge, U.K.: Cambridge University Press.
- Taylor, F. W. (1911). *The Principles of Scientific Management*. New York: Norton & Company.



- Truxler, R., Roth, E., Scott R., Smith, S. and Wampler, J. (2012). Designing collaborative automated planners for agile adaptation to dynamic change. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Vicente, K. J. (1999). *Cognitive Work Analysis*. Mahwah, NJ: Erlbaum.
- Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors*, 44, 62–78.
- Vicente, K. J. and Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-22, 589–606.
- Vicente, K. J., Roth, E. M. and Mumaw, R. J. (2001). How do operators monitor a complex, dynamic work domain? The impact of control room technology. *International Journal of Human Computer Studies*, 54, 831–856. Available online at: <http://www.idealibrary.com>.
- Wall, P. and Brun-Cottan, F. (1995). Using video to re-present the user. *Communications of the ACM*, 38(5), 61–70.
- Wall, P. and Mosher, A. (27–28 October 1994). Representations of work: Bringing designers and users together. In *PDC'94 Proceedings of the Third Biennial Conference on Participatory Design*, Chapel Hill, NC (pp. 87–98). Palo Alto, CA: Computer Professionals for Social Responsibility.
- Watson, M. O. and Sanderson, P. (2007). Designing for attention with sound: Challenges and extensions to ecological interface design. *Human Factors*, 49(2), 331–346.
- Watts-Perotti, J., Sprague, M. A., Swenton-Wall, P., McCorkindale, C., Purvis, L. and McLaughlin, G. (2011). Exploring documents and the future of work. In P. Szymanski and J. Whalen (Eds.), *Making Work Visible: Ethnographically Grounded Case Studies of Work Practice*. Cambridge, U.K.: Cambridge University Press.
- Watts-Perotti, J. and Woods, D. D. (2007). How anomaly response is distributed across functionally distinct teams in space shuttle mission control. *Journal of Cognitive Engineering and Decision Making*, 1(4), 405–433.
- Woods, D. D. (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In G. A. Klein, J. Orasanu, R. Calderwood and C. E. Zsombok (Eds.), *Decision-Making in Action: Models and Methods* (pp. 228–251). Norwood, NJ: Ablex Publishers.
- Woods, D. D. and Dekker, S. (2000). Anticipating the effects of technological change: A new era of dynamics for human factors. *Theoretical Issues in Ergonomics Science*, 1, 272–282.
- Woods, D. D. and Hollnagel, E. (1987). Mapping cognitive demands in complex problem-solving worlds. *International Journal of Man-Machine Studies*, 26, 257–275.
- Woods, D. D. and Hollnagel, E. (2006). *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. Boca Raton, FL: Taylor & Francis Group.
- Woods, D. D. and Roth, E. M. (1988). Cognitive engineering: Human problem-solving with tools. *Human Factors*, 30, 415–430.
- Woolfe, G., Perotti, J. and Sprague, M. (2011). *Method and system to determine preferred color management settings for printing documents*. U.S. Patent #7,894,111. Washington, DC: U.S. Patent and Trademark Office.
- Xiao, Y. (2005). Artefacts and collaborative work in healthcare: Methodological, theoretical, and technological implications of the tangible. *Journal of Biomedical Informatics*, 38, 26–33.
- Zuboff, S. (1987). *In the Age of the Smart Machine*. New York: Basic Books.

---

# 4 Methods for Collecting and Observing Participant Responses

*Sarah Sharples and Sue Cobb*

## CONTENTS

|                                                                            |     |
|----------------------------------------------------------------------------|-----|
| Introduction.....                                                          | 84  |
| Types of Self-Report Methods .....                                         | 84  |
| Stages of a Research Study .....                                           | 85  |
| Selecting a Sample .....                                                   | 85  |
| Biases in Data Collection and Responses .....                              | 86  |
| Study Preparation and Ethics .....                                         | 87  |
| Structured Reporting: Interviews, Diaries and Observations.....            | 87  |
| Interviews and Focus Groups.....                                           | 87  |
| Interviews .....                                                           | 88  |
| Focus Group .....                                                          | 88  |
| Diaries .....                                                              | 89  |
| Structured Observations .....                                              | 91  |
| Measuring Perceptions: Psychophysical Methods.....                         | 93  |
| Measuring Psychological Intensity .....                                    | 94  |
| Accuracy of Judgements .....                                               | 94  |
| Characteristics of Human Judgements: Weber's Law .....                     | 95  |
| Fechner's Law .....                                                        | 95  |
| Method of Limits.....                                                      | 95  |
| Method of Adjustment.....                                                  | 96  |
| Method of Constant Stimuli.....                                            | 97  |
| Characteristics of Human Judgement and Problems in Ensuring Accuracy ..... | 98  |
| Eliciting Comparative Judgements .....                                     | 98  |
| Ranking .....                                                              | 99  |
| Paired Comparisons .....                                                   | 100 |
| Measuring Opinions and Attitudes: Rating and Scaling Techniques.....       | 102 |
| Simple Rating Scales.....                                                  | 102 |
| Techniques for Generating Rating Scales .....                              | 104 |
| Thurstone's Method of Equal-Appearing Intervals.....                       | 104 |
| Likert Scale Development Analysis .....                                    | 104 |
| Questionnaires.....                                                        | 108 |
| Instructions and Background Information .....                              | 109 |
| Questionnaire Format.....                                                  | 109 |
| Question Content.....                                                      | 110 |
| Question Wording.....                                                      | 110 |
| Practical Issues .....                                                     | 111 |

|                                                                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| General Considerations in Collecting and Observing Participant Responses.....                                                                         | 111 |
| Conclusions.....                                                                                                                                      | 112 |
| Appendix: Application for Approval of Research Study Involving Human Participants Used<br>in the University of Nottingham Faculty of Engineering..... | 113 |
| References.....                                                                                                                                       | 116 |

## INTRODUCTION

In many contexts of ergonomics/human factors (E/HF) research, we collect information from people and about people in order to inform our understanding of workplace tasks, contexts, products and equipment and support redesign. The challenge of how to collect such information, balancing considerations of intrusion, validity, detail and data type, is one that has provoked much debate within the field of E/HF (see, e.g. Annett, 2002; Hignett, 2014; Kanis, 2014, Kanis et al., 2014; Stanton, 2014; Vink, 2014). In many situations, both in laboratory and field studies, one method that we commonly use to understand how people feel about a workplace, product or system, or why they are completing a work task in the manner we observe, is to ask them. This process of ‘asking’ may take the form of an interview or informal conversation, but may also have a structured element that perhaps allows us to report their response in a quantitative format or imposes a verbal or descriptive structure on the way in which they respond. The rationale behind moving beyond an open conversation or discussion is that by structuring the way in which we ask questions, we start to be able to understand the generalisability of any findings and are also able to summarise our findings for incorporation into other ergonomics tools or reports.

Such structured tools are normally described as ‘self-report’ methods, denoting the role of the user or operator in providing the information. They are also frequently classified as ‘subjective methods’ – denoting the inevitable subjectivity that will influence the reports of experiences, states and feelings by an individual, through the need to interpret their own feelings and report them in the manner required. Of course, one of the goals of so-called subjective methods may in fact be to reduce the impact of this subjectivity on the outputs from the tools applied.

This chapter describes some different types of tools that an E/HF practitioner may consider using and presents considerations that those using such tools should address whilst using them to inform design requirement specifications and evaluations.

## TYPES OF SELF-REPORT METHODS

Self-report methods often consider the psychological aspects of work. Psychological measurement can be broken down into three types: *psychophysical*, *scaling* and *psychophysiological*. Psychophysical methods refer to a specific class of measures that quantify the relationship between a physical characteristic of the work environment or system (e.g. noise level, temperature, seating design) and the human sensing of that characteristic (e.g. perceived volume, warmth or comfort). Scaling can be used to elicit perceptions of such senses, such as ratings of comfort, but can also be used to report on constructs that are related to the combined effects of a set of different characteristics of a work system (rather than a single physical characteristic as is usually considered in psychophysics), such as rating of job satisfaction, workload or situation awareness (see Chapters 18 and 19). Psychophysiological methods present the biological reaction to a psychological stimulus (e.g. an increase in heart rate that might occur as a result of experiencing stress). In addition to these methods, which all attempt to quantify the experience of an individual to some extent, there is a fourth class of method, termed here *structured reporting*, which covers methods including structured group and individual interviews, verbal protocols, open-ended questionnaire items and longitudinal methods such as diaries.

This chapter covers structured reporting, psychophysics and scaling. Particular approaches that we focus on here are interviews, focus groups, diaries, observation, psychophysical methods, ratings, scaling techniques and questionnaires. It is important to note that several other texts cover each of these topics in more depth than is possible in this short chapter, and the reader is referred to specialist texts such as Robson (2002), Blake and Sekuler (2006) and Oppenheim (1992) for a more in-depth discussion of some of the concepts presented here. Psychophysiological methods are covered in Chapters 21 and 22.

## STAGES OF A RESEARCH STUDY

Before considering the details of individual methods that should be applied, it is worth outlining the different stages that should be undertaken when conducting a research study. Some of the methods outlined in this chapter may be administered as part of other types of E/HF activities, such as elicitation of requirements for workplace redesign or a study designed to compare task performance and/or participant opinion under different experimental conditions (see Chapter 2), and the word ‘research’ should not be assumed to apply only to data collected in an academic setting – many practitioner activities also require the collection of data from participants using self-report or subjective methods, and so would follow some or all of the stages outlined here.

The main stages of an E/HF study are

- Choosing a research setting
- Selecting variables for consideration and establishing research questions
- Identifying participant sample characteristics
- Deciding on the data collection method
- Checking if data collection meets requirements for data analysis and reporting
- Collecting data
- Analysing, interpreting and presenting results

Books such as Dane (1990), Robson (2002), Field and Hole (2003) cover these topics in excellent detail, but a few issues pertinent to E/HF investigations will be considered here, before presentation of specific methods for self-report and subjective data collection in depth.

## SELECTING A SAMPLE

The primary mantra when identifying an appropriate set of participants is to ensure that the sample is representative of the population to which you want to generalise *in terms of the characteristics under investigation*. A study needs to consider *who* to select, *how* to select them and *how many* to select. Therefore, if a study is related to accessibility of an interface for older adults, the sample must consist of older adults; if however the study simply requires participants with a range of anthropometric characteristics to participate, the attitudes towards technology of the participant sample may not be important. Even with a well-designed study, if a participant sample is not carefully recruited, then a source of bias may be introduced. For example, if an advertisement is placed to recruit participants, then inevitably those who are interested in or willing to participate in research studies are more likely to respond; the presence of incentives such as expenses payment may also bias the composition of the participant sample.

Finally, the number of participants required for a study is a matter of much debate. Some statistical techniques (e.g. chi-square) specify a minimum number of participants before the data from them can be analysed (at least five participants in each ‘expected’ cell of the analytical table for a two-way analysis), but in other cases for statistical analysis, it is important to understand the statistical ‘power’ of the sample size and analysis method used (see Chapter 2 for further discussion). An approach that is derived from qualitative analysis, but is useful to consider as an underlying

principle in all E/HF investigations, is that of ‘data saturation’ – continuing to collect data until no further issues or ideas are derived. This is related to the ‘five-user’ principle that is often discussed in the field of usability, where it has been suggested that five experts will elicit 75% of usability problems with an interface (Nielsen, 2000). Overall, sufficient data should be obtained to reliably assess the effects of variables or answer the questions under investigation with minimum cost in time and resources (see also Nielsen, 2012).

## BIASES IN DATA COLLECTION AND RESPONSES

In addition to the selection of participants, the design of the methods to elicit subjective responses from participants may themselves lead to bias in data if not considered carefully.

The first type of bias that might be encountered occurs as a result of the study size or duration. If a study is long or repetitive, participants can exhibit fatigue effects, where performance worsens over time, or where they become annoyed with the time taken to complete a study. Fatigue effects can be seen, for example, with overlong questionnaires; therefore, good practice is to clearly inform the participant at the start of the survey how long it is likely to take to complete the questionnaire and, in the case of surveys delivered using online tools, provide an indicator of how far through the questionnaire a participant is (this information can of course also be gleaned from progress through a paper-based questionnaire). Training effects may also result from a repetitive study, especially if there is any element of ‘skill’ involved in task completion; therefore, either the study or task should be designed to minimise such effects (e.g. by only selecting expert participants) or a training element may be included in a study, ideally to ensure that participant performance reaches a measured ‘plateau’ before the active data collection part of the study begins.

The second type of bias that might be encountered is associated with the participant either intentionally or unintentionally altering their response as a result of participating in the study. This may come into the category of the Hawthorne effect where participants respond positively simply because they are aware that they are being observed (see later in this chapter for a more detailed description of this effect). Participant bias may also be due to the participant either ‘guessing’ or being informed about the purpose of the investigation, and altering their behaviour as a result. Examples of this type of bias include beneficent and maleficent participant bias, where a participant believes that they are aware of the experimental hypothesis, and either try to support (beneficent) or disprove (maleficent) the hypothesis as a result. An alternative type of participant bias that we have termed the ‘macho effect’ is where participants (both male and female) become overly competitive whilst participating in a study and try to work hard to mask any negative effects they might be experiencing, such as fatigue effects as a result of completing a strenuous task. The use of high-resolution performance measurement tools and the removal of any competitive ‘incentives’ (which might include feedback on performance during the trial) can minimise such bias.

The final type of bias is that associated with the investigator. With the best will in the world, if an investigator has an implicit or explicit hypothesis or theory before a study, their desire to obtain results that confirm their expectations may affect their behaviour whilst collecting the data. Within the field of clinical trials, this issue is addressed via the use of double-blind techniques; such sophisticated experimental designs are less frequently used within E/HF, but efforts should still be made to reduce investigator bias in all studies. This is best achieved by the use of clear structured protocols – for example to avoid the use of more positive responses to interview comments that agree with the investigator’s view, the investigator could pre-plan responses and practise interviewing with colleagues before they interview participants; they could also ensure that they explain the purpose of the study using written material, so that there is no unintentional variation in the way in which they communicate their study in spoken communication. Ultimately, the avoidance of such bias, such as varied intonation in the voice in response to participant behaviours or comments, relies on professionalism and expertise of the investigator, and if this type of bias is of concern then, if possible, a naïve investigator should be asked to conduct data collection.

## STUDY PREPARATION AND ETHICS

When conducting any study or observation or collecting data about individuals, it is essential that full consideration is given to ethical issues and that steps are taken to ensure participant well-being throughout the study. All research institutions are required to adhere to the principles of good practice in conducting research as specified by relevant governing bodies or societies (examples in the United Kingdom include the Institute for Ergonomics and Human Factors, the British Psychological Society, the British Educational Research Association, and the Economic and Social Research Council). Before planning your research study, you should be aware of the relevant research regulations and governance bodies applicable to your area of study, and you should ensure that you follow the required procedures of your institution for ethical approval.

It is usual for ethics submission to require the use of a pro forma detailing the specific information required by the review committee, and these may differ for ethical committees in different institutions or disciplines. However, the researcher should ensure that they can answer all possible questions that a study participant may ask. The following list may be useful as a general guide.

Participants involved in research studies have a right to

- Know the goals of the study and who is funding the work
- Make an informed decision about whether or not they wish to participate
- Leave the study at any time if they do not wish to continue
- Know what will happen to them during the study and how long it will take
- Know if they may experience any discomfort
- Know what will happen to the findings
- Privacy of personal information
- Be treated courteously

Normally, this information should be given in a participant information sheet at the beginning of the study, and participants should be required to confirm that they have understood the nature of the study and that they are happy to participate. *All* participants must be told that any involvement in the study is voluntary, and they are free to withdraw at any time. You should also explain any consequences for the participant of withdrawing from the study and indicate what will be done with the participant's data if they withdraw.

Good practice is to include all of this information in submission for ethical approval. As an example, a modified copy of the ethics application form used in the Faculty of Engineering at the University of Nottingham is given in the Appendix.

The following sections describe the specific considerations associated with the different types of methods discussed in this chapter: structured reporting (interviews, diaries and observations), psychophysics, eliciting comparative judgements and measuring opinions and attitudes. Finally, reflections on gathering a range of these approaches together in a questionnaire are also presented.

## STRUCTURED REPORTING: INTERVIEWS, DIARIES AND OBSERVATIONS

### INTERVIEWS AND FOCUS GROUPS

A common method of eliciting subjective and self-report data from participants is the application of investigator-led questioning, in the form of interviews and focus groups. Interviews are normally conducted with individuals and usually follow a structured or semi-structured protocol. They can be used to ask questions about facts, behaviours, beliefs and attitudes, and may include the application of formal techniques such as repertory grid or critical decision method (see Chapters 7 and 3). Interviews and focus groups are often very valuable for establishing issues that can then be evaluated using quantitative measures – for example in an examination of the thermal environment,



an investigator may conduct interviews to find out the areas of a work environment in which problems are most frequently experienced, to direct where instrumentation should be used to collect more detailed data.

## Interviews

Advice for conducting an effective interview includes providing a clear introduction, explaining who you are and how data from the interview will be stored, analysed and reported; ensuring that questions are phrased in an understandable and non-threatening manner; ensuring that the interviewee (and not the interviewer) is the person who does more of the talking; avoiding leading questions and carefully considering the use of specialist jargon in questions; avoiding long or double-barrelled questions (e.g. what do you feel about the current technology you use compared with the technology you were using 5 years ago?); and identifying pre-planned probes or prompts that you will use if an interviewee does not initially answer your question fully. Usually in interviews, you would want to capture as much detail as possible. One way to ensure this is to audio-record the session (however, you must ensure that the intention to record is noted in your ethics application and that your participant has given consent for their responses to be recorded).

Advantages of interviews include their flexible and adaptable nature, providing the opportunity for an expert interviewer to modify the line of enquiry in response to interviewee comments. If interviews are face to face, they also provide the opportunity to capture non-verbal cues (although these can be ambiguous to interpret). Detailed linguistic analysis (e.g. considering pauses and phraseology as indicators of bias or attitude) may also be applied if the analyst has sufficient specialist knowledge. The data from interviews can be very rich in detail and, as described in Chapter 5, can provide the basis of a range of different qualitative analysis approaches.

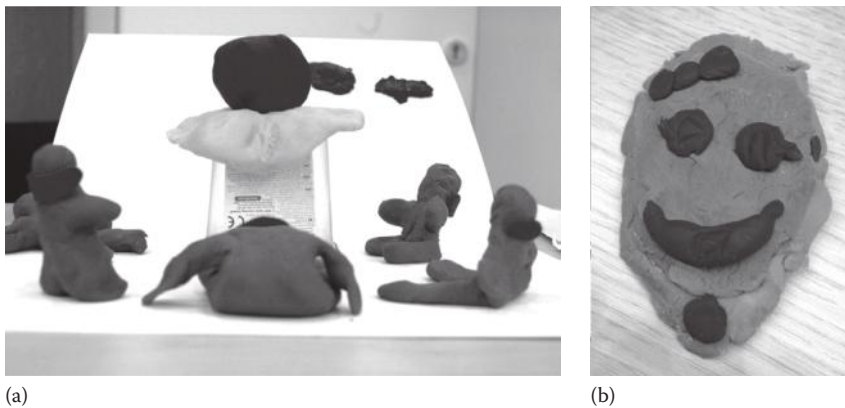
To be effective, interviews must be carefully prepared and, as qualitative analysis will be applied, can be time consuming (particularly at the analysis stage). They are often inevitably subjective, and indeed, some disciplinary perspectives, such as the application of ethnographically informed techniques, explicitly acknowledge this. It is therefore very important in the context of E/HF to be clear about the subjectivity involved in the use and reporting of interviews; whilst in an academic paper it may be quite clear what the perspective of the researcher is, and therefore how the data should be interpreted, if data are being reported into an industrial context, they may be summarised or reviewed, and the ‘auditability’ of the origin of the data and its interpretations may be lost. Therefore, this needs to be taken into account when using subjective approaches to interviews within E/HF assessments and interventions.

## Focus Group

Focus groups are often used as an efficient method of obtaining a range of views on a work context or new design intervention. A focus group can simply be considered to be a group interview and thus can be conducted with any number of participants, from 2 to around 10. When the number of participants in the group is small, it can often run similarly to an interview, with care required by the interviewer to ensure that all participants are given the opportunity to respond to questions and that no one individual dominates. When the numbers are greater than around four people, more formal approaches can be adopted to ensure that individual views are represented. It may or may not be appropriate to reach consensus – normally within a focus group, it is acceptable to obtain a range of opinions, and there is no expectation of consensus; however, in some situations, such as participatory ergonomics (see Chapter 34), the value of consensus can be higher.

Formal methods to ensure that all views are represented can include asking participants to first write down their views before discussing them, using collaborative spaces such as flip charts to capture written comments or using modelling clay or drawing materials to encourage people to express their views or ideas in a more ‘fun’ format. Figure 4.1 shows two examples of clay models produced (1) by a group working together and (2) by a child. In both cases, the participants were asked to use the modelling clay to communicate their visions of future technology.





**FIGURE 4.1** Examples of modelling clay representations of participant visions of (a) future technology to support design practice and (b) a future TV remote control concept.

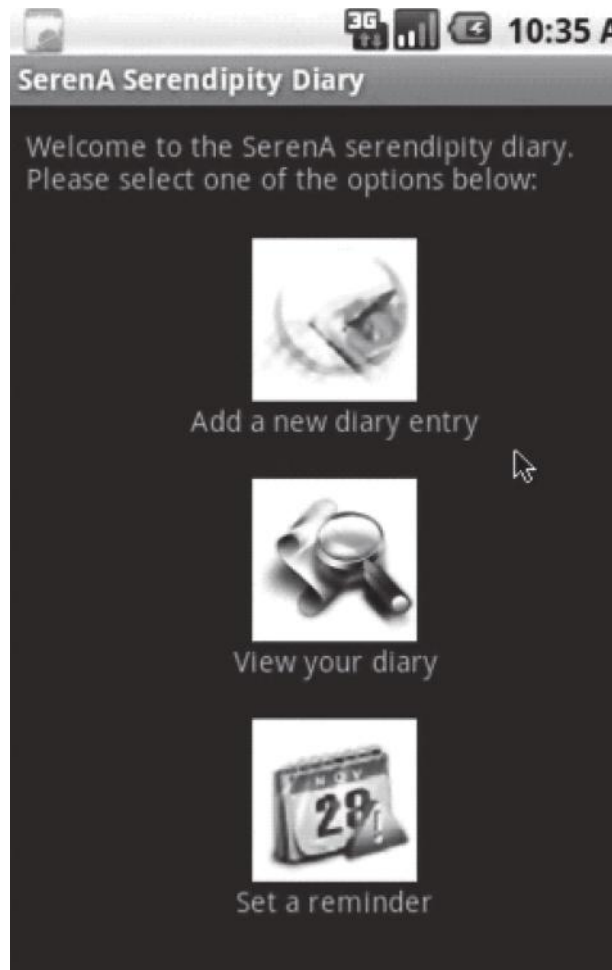
The value of focus groups is in the ideas that an individual respondent might prompt and the immediate feedback given by the group on this idea, or other ideas or thoughts it might propagate. There is considerable skill to facilitating an effective focus group. The room in which the focus group is to be held should be comfortable and private, and the participants should be made to feel welcome. It is important to emphasise that ‘all are equal’ in the focus group context, and it may in fact be appropriate to, for example, hold different focus groups with people with different seniorities of job roles within an organisation, to avoid the dominance of more senior employees within the group. The facilitator should ensure that they manage the conversation, politely moving on from any dominant individuals or off-topic conversations, and particularly encouraging those more reluctant to speak to participate. Simple things such as ensuring that all participants wear name badges so that they can be addressed by their preferred name can be very effective in supporting the social atmosphere within the group. As with other methods presented in this chapter, the aims of the focus group should be clear, and the E/HF expert should make sure that all participants are aware of how their data are going to be reported and analysed.

## DIARIES

A type of data collection that is currently increasing in popularity is the use of diary studies. Before the prevalence of portable tools to facilitate diary studies, such as smartphones or tablet computers, diary studies were known to be useful but difficult for participants to adhere to, with very high participant dropout rates during a study. However, the development of a range of tools to support diary data has led to a recent resurgence in interest in such approaches.

The advantages of a diary approach are that the experiences of participants can be tracked over time. This means that a range of different circumstances can be captured (e.g. times of day, levels of activity, environmental conditions) and also that it has the potential to reduce the biases that may result from the presence of an experimenter. Figure 4.2 shows a screenshot of a diary application used in a project that examined the potential use of technology to support serendipity in research and scientific discovery.

Cranwell et al. (2012) and Sun et al. (2013) report a set of studies that implemented the diary approach using smartphone tools. These studies, which examined journey behaviour, energy use and research practice, demonstrated how technology could be used to support not only written diary entries but also voice notes, video and pictures. In addition, if relevant, the device could also be set to record the participant’s location, and such technologies, possibly with peripheral sensors added, will also have the potential to record environmental characteristics (e.g. background noise,



**FIGURE 4.2** Example of application used to elicit experiences of serendipity from participants.

temperature, humidity). Online diaries are therefore a promising tool that E/HF specialists should be aware of over the next few years.

Our work has enabled some initial insight into the value of such tools for a range of types of investigation. We found that an open and exploratory tool, where the respondents may not be expected to respond at a particular time of day or in a particular circumstance (e.g. when coming up with new design ideas for an office context), can be particularly useful at the early stages of a design process. The use of diaries can overcome some of the challenges associated with more overt methods such as focus groups, allowing those who may be less confident to share their ideas to participate (although people may also be reluctant to share ideas using a digital medium if they are not clear how their data will be stored or analysed). More structured studies, where participants are perhaps prompted for a response at a specific time of day, can be helpful in hypothesis testing contexts. For example, Cebola et al. (2012) used an online diary approach to capture experiences of on-call work in the U.K. rail industry; this implementation of the diary had a particular advantage in that it used the devices that respondents were already using as part of their work activities.

When implementing diaries, the use of prompts or reminders should be carefully considered; too many prompts can quickly become annoying and will reduce participation in the study. Our experience shows that if the study is interesting or engaging, participants are willing to receive up to three prompts or reminders a day, but this will only be sustained over a short period (e.g. 1 week)

and is normally accompanied by an incentive payment to the participant (which can present ethical concerns in some circumstances).

As with other more qualitative methods, the analysis of diary study data should be carefully planned and can be time consuming. We have found that diaries are particularly useful as memory aids to support more in-depth questioning – if a participant has several notable experiences relevant to the research question, then they use the diary to quickly record these instances but, in a follow-up interview, describe the issues associated with this instance in more detail. This combination of a diary and interview method has proved particularly fruitful.

It may seem obvious, but the use of diaries is particularly useful when the participant is mobile and physically remote from the E/HF investigator. Combining records such as photos or text entries with position tracking information can yield very rich data, but it is important to ensure that the participant is aware of when and how their movements are being tracked, and for them to easily be able to switch off any remote tracking when desired.

The final point to note with diary approaches, particularly when using tools such as smartphones to deliver the tools, is that they can be a fun and engaging way of collecting user data. Participants feel that effort has been invested in them, that the means of obtaining the data has been carefully thought about and made as easy as possible for them and that the use of the device might even prompt conversation amongst their friends and work colleagues, which may make them think further about their input into the investigation. We are fortunate in E/HF domains in that frequently we are working in partnership with our participants, as opposed to treating them as ‘subjects under investigation’; therefore such engagement, although it may bring a small amount of bias, is normally a good thing that increases the richness of data obtained and the satisfaction of the participant with their involvement in the study.

## STRUCTURED OBSERVATIONS

Whilst the method of observation is covered to some extent in Chapters 3, 5 and 6, an approach that is often applied but is not covered elsewhere is that of ‘structured observation’. When structured observation is applied, a researcher has a clear research question or hypothesis which they wish to test, and rather than in an ethnographic approach where they aim to build a rich picture of the environment, they collect information based on a pre-defined set of categories or framework.

Such observation can be either *direct* or *indirect*. Direct observation involves the researcher being present in the immediate location or using remote observation techniques such as closed-circuit television (CCTV). Indirect observations are taken where the task is viewed at another time from when it took place by other means such as video recording.

Drury (1995) notes the value of direct observation, ‘Merely walking around an operating human/machine system to “see what is going on” is hardly an observational method worthy of the name, yet it is an indispensable first step...’. Observation allows us to capture information such as occurrences and timings of events, sequence of operations (including errors), behaviours and interactions. Sharples et al. (2011) present an example of an application of an observation framework to analyse interactions by rail signallers with a range of different levels of automation in rail contexts (see Table 4.1 for a simplified version of the output from this study). The example data in the table represent the percentage of time spent on each type of activity over a set period, with observed tasks being sampled every 5 s for 48 participants over periods of between 60 and 90 min.

When observing a workplace, there are a number of different roles that the investigator can take. These are

1. Complete participant (‘going native’ to the extent that other participants may not be aware that the investigator is conducting a study)
2. Participant as observer (the participants are aware of the observer’s role)

**TABLE 4.1****Example of Quantitative Structured Observation Data Collected in Rail Signalling Context (Mean and Standard Deviation of Percentage Time Distribution between Activity Types)**

|                | Entry–Exit Panel (NX) | VDU without Automation | VDU with Automation |
|----------------|-----------------------|------------------------|---------------------|
| Planning       | 13.30% (8.37%)        | 8.33% (6.14%)          | 11.14% (6.14%)      |
| Monitoring     | 56.94% (13.8%)        | 69.49% (10.53%)        | 48.16% (15.68%)     |
| Interaction    | 17.35% (7.3%)         | 11.78% (5.13%)         | 9.04% (4.47%)       |
| Quiet time     | 3.54% (3.52%)         | 4.38% (5.33%)          | 22.38% (18%)        |
| Communications | 8.86% (5.96%)         | 5.34% (4.38%)          | 7.78% (3.69%)       |
| CCTV           | 0% (0%)               | 0.69% (1.27%)          | 1.50% (3.21%)       |

*Source:* Simplified from Sharples, S. et al., Impact of automation, 2011.

3. Marginal participant (where the observer may, e.g., play a small part in an activity)
4. Observer as participant (the group being observed is expanded to include the role of researcher)
5. Pure observer (the observer purely observes the action)

It can be seen that these form a sort of ‘continuum’, and there are advantages and disadvantages to each approach. For example, in the case of the first approach, it is important that ethical concerns are addressed and that the safety and integrity of the observer is preserved.

It is known that when people are aware that they are being observed, they may not behave in their usual manner whilst they are being watched or when being interviewed about carrying out an activity. The phenomenon was first documented by researchers in the late 1920s in a project looking at physical and environmental influences of the workplace, such as lighting and heating, on production in the Hawthorne Plant of the Western Electric Company in Cicero, Illinois. The major finding of the study was that almost regardless of the experimental manipulation employed, the production of the workers seemed to improve. One reasonable conclusion is that the workers were pleased to receive attention from the researchers who expressed an interest in them and that it was this attention, rather than the experimental manipulation, that produced the measured change in performance. This phenomenon is now known as the ‘Hawthorne effect’. The original study was expected to last only 1 year, but because the researchers were set back each time they tried to relate the manipulated physical conditions to the worker’s efficiency, the project extended out to 5 years. It is necessary to consider whether such effects may be present and to try and control for this (Fostervold et al., 2001). E/HF investigators need to try and be as unobtrusive as possible when collecting data if taking on a ‘pure observer’ role.

Sometimes the opposite approach can be taken, in which it is explicitly acknowledged that an observation or analysis is taking place and that participant behaviour is likely to be affected by this. In such cases, the E/HF practitioner may consider the use of verbal protocol analysis (see Chapter 7 and Bainbridge and Sanderson (2005) in a previous volume of this book) where the participant explicitly verbalises their thoughts and strategies whilst completing the task. Although it is inevitable that this introspective method will change the nature of the interaction, the richness of the data obtained from the verbalisations, coupled with observed interaction, can enable strong inferences to be made about cognitive activities.

When carrying out an observation study, the E/HF practitioner should consider a number of issues, including data capture, recording methods and the method of observation itself. In terms of data capture, the main consideration is whether to use ‘people’ as the conduit for capture or technology-led solutions. Human observers have the advantage that they are able to ‘capture the unexpected’, and if the observer has sufficient expertise, they can be conducting analysis whilst observing.

Of course, humans suffer from the usual problems of fatigue and loss of attention, and the presence of the human observer can be off-putting to participants. Increasingly, the use of hardware (e.g. activity logging, sensors present in workplace) is becoming feasible as a method for capturing workplace performance and interaction. This offers rapid collection of large data sets and removes any subjectivity from data collection and interpretation (although of course there is still some subjectivity in the selection of data to be captured in the first place). This approach also does not suffer from fatigue or attention deficit effects, and can be cost-effective. However, the volume of data can rapidly become vast, and the analysis and interpretation of such data can be time consuming. We are currently applying such techniques in a hospital setting, tracking movement of hospital employees as they complete their work and integrating these data with performance data recorded from a mobile-based job allocation system. The analysis of these data requires considerable effort to set up an analysis model to support data extraction and coding; once this model is established, we estimate that 1 h of data from tracked movement and task recording (both observed and logged) would take around 6 h to analyse.

Advantages of an observation approach are that the data have high face validity, they provide excellent case study examples and give real-life insight into workplace interaction that can be used to complement other, perhaps laboratory-based, data. However, although it can be difficult to generalise from the data, it is an approach that tends to be resource intensive, and the presence of an observer, whether direct or indirect, can affect the actions being observed. Finally, the value of observation – its rich picture of a real workplace context – can mean that caution should be applied if generalising findings from one workplace context to another.

MEASURING PERCEPTIONS: PSYCHOPHYSICAL METHODS

Psychophysics is concerned with the relationship between physical stimuli present in the environment and perceived sensations (or psychological responses) experienced by individuals in that environment. It is extensively used in an examination of visual and auditory capabilities, and in E/HF, research has value in exploring perception of the environment in terms of factors affecting comfort and performance, such as temperature, contact pressure or weight. Table 4.2 shows some examples of physical stimuli that can be inferred from perceived sensations or responses using psychophysics techniques.

Psychophysics techniques aim to quantify sensory experiences in a manner that allows further understanding of the relationship between the physical environment and how it is perceived by an individual. Measurement of the strength or size of the physical stimulus is usually straightforward – we can record the level of magnitude along the appropriate measurement scale (e.g. degrees Fahrenheit/Centigrade, pounds/grams and decibels) using specialist instrumentation and sensors. Psychophysical techniques are based upon the notion that it is difficult for people to accurately judge an absolute rating of intensity of a sensation from participants, but that people are generally more

**TABLE 4.2**  
**Physical Stimuli and Associated Perceived Sensations Measured**  
**Using Psychophysics Techniques**

| Physical Stimuli (S) | Perceived Sensation or Psychological Responses (R) |
|----------------------|----------------------------------------------------|
| Wavelength of light  | Colour                                             |
| Frequency of sound   | Pitch                                              |
| Temperature          | Sensation of warmth/coldness                       |
| Contact pressure     | Discomfort                                         |
| Weight               | Perceived heaviness                                |

reliable at giving relative judgements of perception of signal strengths. Gustav Fechner (1860) proposed methods for measuring perceived sensations based on subject comparisons of different levels of sensory intensity of the physical stimulus.

In the context of E/HF, psychophysical measures are most useful for

1. Detecting thresholds: the boundary between a stimulus giving no sensation and that giving a slight sensation (e.g. a sound which is just audible)
2. Detecting differences: the smallest change reliably detected (e.g. discrimination between colours)
3. Judgements of the direction and magnitude of differences: the accuracy of judgements (e.g. levels of comfort of different seats)

This section provides a brief introduction to some of the considerations and approaches required when applying psychophysics methods. For more in-depth description of the methods, see texts such as Gescheider (1985, 1997) and Blake and Sekuler (2006). Examples of the application of psychophysics in E/HF can be found in the *Ergonomics* journal papers (e.g. Garneau and Parkinson, 2013; Maikala et al., 2014).

## MEASURING PSYCHOLOGICAL INTENSITY

Fechner (1860) suggested that whilst perceived sensations cannot be directly measured against the sensory intensity of the stimulus, they can be measured against each other. That is, humans are able to detect two sensations and make a judgement as to whether they are the same or different.

Two types of sensory thresholds are measured:

1. The *absolute threshold* (originally termed Reiz Limen (RL)): The lowest level of stimulus intensity required for a participant to perceive it, that is, the weakest level of a stimulus that is *just detectable*.
2. The *difference threshold* (originally termed Differenz Limen (DL)): As the sensory intensity level of the stimulus changes (increases), participants are asked to state when the sensation they experience is *just greater* than the previous sensation. This may also be referred to as the *just-noticeable difference* (jnd) and describes the participant's ability to discriminate between different levels of sensory intensity of the stimulus.

## ACCURACY OF JUDGEMENTS

All psychophysics methods rely on simple judgements of equality, that is, 'greater than' or 'less than'. Humans are good at making such judgements of equality, but not as good at making judgements in magnitude or size of a difference. There is an element of uncertainty in making any such judgements, especially when close to a threshold. Thresholds may not remain stable even for an individual. Thresholds and small differences therefore have to be defined in terms of the probability with which the comparison stimulus (Co) is judged different from a standard (St), and this is calculated from a series of repeated measurements. This leads to the definition of the *absolute threshold* as 'the value which evokes a response 50% of the time' and the *difference threshold* as 'the difference between two stimuli such that the participant will report that there is a difference on 50% of occasions'.

Judgements of equality are not always accurate and may also be affected by the surrounding environment or by the test conditions. A constant error (positive or negative) may occur. A *point of subjective equality* (PSE) is therefore defined in statistical terms as 'the magnitude of the stimulus at which a comparison of stimuli is most likely to result in a judgement of "equal" or "same"'.



### CHARACTERISTICS OF HUMAN JUDGEMENTS: WEBER'S LAW

The DL changes in magnitude in relation to the distance from the initial stimulus value. For example, if the luminance of a light was being changed, it may be the case that a participant is able to perceive the difference between a light of 100 and 110 units – that is, the DL is 10 units. However, if the lower luminance is changed to 1000 units, then the person may not notice a difference until a much larger change had taken place, for example up to 1100 units (i.e. 100 units difference).

Weber's law shows that the DL is not constant, but that it is a constant *proportion* of the initial stimulus value. For example, in the lighting example given earlier, the DLs were both 10% of the initial stimulus.

This can be represented as

$$\frac{\text{Difference threshold}}{S} = K$$

where

DL is the amount by which the stimulus intensity must be changed in order to produce a just noticeable change

S is the stimulus intensity which was started with

K is a constant proportion

In fact, K remains fairly constant throughout the middle range of intensity of a sense, but varies widely from one sense to another (see Gescheider, 1997). Examples of K are light intensity (K = 0.02), sound intensity (K = 0.04) and lifted weight (K = 0.02) (Teghtsoonian, 1971).

### FECHNER'S LAW

Responses to physical stimuli are non-linear (e.g. decibel scale). This means that large differences in stimulus intensity (e.g. sound pressure level) correspond to relatively small changes in perceived stimulus (e.g. loudness). However, Fechner showed that there is a linear relationship between the response R and the logarithm of the stimulus S.

$$R = c \log S \text{ (where } c \text{ is a constant)}$$

Three main psychophysical methods that are used for E/HF studies are *method of limits*, *method of adjustment* and *method of constant stimuli*. Each of these is described in the following text.

### METHOD OF LIMITS

Also known as the method of minimal changes, method of serial exploration or method of least noticeable differences, the method of limits is used to establish *RL* and *DL* by repeated presentation of a comparison stimulus (*Co*) that approaches and recedes from a standard stimulus (*St*) in short steps. The threshold is established as the average of *jnd* (just noticeable difference) and *jnnd* (just noticeable no difference).

The absolute threshold (*RL*) is determined by presenting the participant with a descending series of a weak stimulus. After each stimulus, the participants respond yes or no to report whether they can perceive the stimulus or not. The trial continues until the stimulus is no longer perceptible. The process is then repeated with the presentation of a series of stimuli in the ascending order of intensity, starting with one that is believed to be not detectable and continuing until the participants respond with a 'yes' to indicate that they can now detect the stimulus.



This process is continued, alternating between series of descending and ascending stimuli, until the experimenter is satisfied with the relative uniformity of threshold values reported by the participant. It is important that the starting stimulus level for each series is varied to avoid the participant selecting a threshold (T) level at the same place (number of steps) by counting rather than stating what they can actually perceive.

To establish a *DL*, the participant is presented with a stimulus (*St*) and a series of ascending and descending comparison stimuli (*Co*). For each value in the series, the participant states whether the *Co* is greater than (+), equal to (=) or less than (–) *St*. In a descending series, threshold points are located as the first change in response from ‘*Co* is greater than *St*’ T(+), and the first response of ‘*Co* is less than *St*’ T(–). In an ascending series, threshold points are located as the first change in response from ‘*Co* is less than *St*’ T(–) and the first response of ‘*Co* is greater than *St*’ T(+). The series of different trials may produce a range of *St* values between T(+) and T(–). Presentation of descending and ascending series continues until the experimenter is satisfied with the uniformity of responses. The average values of T(+) and T(–) over all the series presented can then be calculated. From this information, a variety of measures can be calculated as shown in Table 4.3.

An example of this method in use was given in Southall (1985) to assess the ability of participants to perceive the force exerted to operate a vehicle clutch pedal and to see whether perceived force is dependent upon the pedal resistance. It is widely used in E/HF studies of hearing and vision and to establish workplace dimensions and identification of acceptable limits of manual force (Fischer and Dickerson, 2014), and pushing and pulling (Garg et al., 2014). Fitting trials are a slightly modified version of the method of limits.

## METHOD OF ADJUSTMENT

Also known as the method of equivalent stimuli or method of average error, this method can be used to measure sensory thresholds (e.g. in hearing tests) or to track changing sensory responses (e.g. to measure visibility threshold during a course of dark adaptation).

In this method, the participant is asked to repeatedly adjust a comparison stimulus (*Co*) until it appears equal to a given reference stimulus (*St*). For repeated trials, *Co* should start at varying directions and distances from *St*. The method is most often used to establish variability in perception of equality and is therefore related to *DL*.

Numerous studies conducted in the 1960s–1980s established the application of the method of adjustment (and adapted forms) for the assessment of maximum acceptable weight of lift (MAWL) in manual handling tasks, leading to publication of tables of maximum acceptable weights and forces (Snook, 1978), later revised (Snook and Ciriello, 1991). However, recent research suggests that lifting load limits identified using these methods are sensitive to instructions presented to the participants (Karwowski et al., 2007). An example of this method in use is given in Li et al. (2007)

**TABLE 4.3**  
**Calculations Made Using Parameters Obtained through Method of Limits**

| Measure                            | Calculation                                       |
|------------------------------------|---------------------------------------------------|
| Interval of uncertainty (IU)       | Average T(+) – average T(–)                       |
| Difference threshold (DL)          | $IU \times 0.5$                                   |
| Point of subjective equality (PSE) | $(\text{Average T(+) + average T(–)}) \times 0.5$ |
| Constant error (CE)                | PSE– <i>St</i>                                    |
| Weber’s fraction (K)               | DL/PSE                                            |

**TABLE 4.4**  
**Calculations Made Using Parameters Obtained through Method of Adjustment**

| Measure                            | Calculation                                   |
|------------------------------------|-----------------------------------------------|
| Point of subjective equality (PSE) | M (i.e. mean of Co)                           |
| Constant error (CE)                | PSE – St                                      |
| Probable error (PE)                | $0.67 \times \text{SD}$                       |
| Average mid-quartile range (Q)     | $(75\%ile - 50\%ile) + (50\%ile - 25\%ile)/2$ |
| Interval of uncertainty (IU)       | $75\%ile - 25\%ile$ or $2 \times \text{PE}$   |
| Weber's fraction K                 | $\text{PE}/\text{M}$ or $\text{Q}/\text{M}$   |

*Note:* PE and Q are equal to the DL of method of limits.

to determine maximal acceptable weight of handling (MAWH) materials under different footwear–floor slipperiness conditions.

To obtain measures for the method of adjustment, the first task is to calculate the mean (M) and standard deviation (SD) of the series of settings of Co. From this information, a variety of measures can be calculated as shown in Table 4.4.

The method of adjustment is economic to use, but there are difficulties in separating motor from perceptual errors and in continuously adjusting some stimuli (e.g. lifting weights).

## METHOD OF CONSTANT STIMULI

Also known as the frequency method, the method of constant stimuli avoids many of the problems or limitations experienced when using the method of limits or the method of adjustment. Examples of this method in use include measures of user perception of depth in stereoscopic displays (e.g. Surdick et al., 1994). However, as mentioned previously, the measures obtained using different psychophysics methods may vary (e.g. Rolland et al., 2002), and so it is important to think carefully about which method is most suitable for the study you are conducting and to review previous research in order to identify potential issues that may impact upon the results.

In this method, the participant is presented with a number (usually 5 or 7) of different levels of the Co stimulus. These should be known to extend beyond the probable levels of DL. A standard (St) is needed in order to measure a difference threshold (DL), but the method can be used without a standard if an absolute threshold (RL) is all that is required.

Ideally, the middle level of the Co stimulus should be predicted to be close to RL (i.e. about the level that could just be detected 50% of the time). The other levels should be spread evenly and with equal stimulus intervals above and below RL. The participant is presented with the different stimulus values in a random order, with about 50–200 repetitions of each level. The participant states yes or no according to whether or not they can detect each stimulus level.

The proportion of yes responses is calculated for each stimulus level. The large number of trials allows for statistical analysis of the results. A variety of methods of analysis of these data are then possible, ranging from simple graphical interpolation to solutions of least-squares fitting.

To analyse data for the DL, the probability scores are converted into Z scores. They are then plotted on a graph of S (x-axis) against Z (y-axis), and a best fit line is then drawn. The values shown in Table 4.5 can be calculated.

*Paired comparison* (see later in this chapter) is a particular example of the method of constant stimuli that does not use a standard, but asks participants to make comparative judgements between each pair of stimuli.

**TABLE 4.5**  
**Calculations Made Using Parameters Obtained through Method of Constant Stimuli**

| Measure                            | Calculation                                                                  |
|------------------------------------|------------------------------------------------------------------------------|
| Point of subjective equality (PSE) | $M = s$ value corresponding to $Z = 0$                                       |
| Standard deviation (SD)            | Difference between $S$ value corresponding to either $Z = \pm 1$ and PSE     |
| Probable error (PE)                | $s$ value corresponding to either $Z = \pm 0.67$                             |
| Constant error (CE)                | $PSE - St$                                                                   |
| Interval of uncertainty (IU)       | 75%ile – 25%ile or difference between points corresponding to $Z = \pm 0.67$ |

*Note:* PE and Q are equal to the DL of method of limits.

### CHARACTERISTICS OF HUMAN JUDGEMENT AND PROBLEMS IN ENSURING ACCURACY

There are various sources of fluctuation and error in psychophysical measurement. These may be constant or variable and are listed as follows:

- *Space effects:* These can be due to the background, geometric arrangement or orientation in the presentation of the stimulus. These are often unpredictable and difficult to explain.
- *Time effects:* The amount of time between presentations of two equal stimuli may influence subjective judgements. For example, the second of two equal stimuli is almost always judged louder, heavier or brighter than the first. However, if two equally intense tones are separated by 0.5 s, the second is usually judged as less loud, but if they are separated by 0.6 s, the second is usually judged louder (Gescheider, 1985).
- *Sequential tendencies:* Series effects may occur in ascending or descending trials. The thresholds may differ because the participant tends to persist in the same kind of response. They may frequently give lower thresholds in descending series than in ascending ones, and vice versa.
- *Expectancy bias:* Participants try to find some basis on which to make consistent decisions, perhaps obtaining cues from the experimenter or equipment. It is therefore important not to provide any knowledge of results during the trial.

The participant's sensitivity may change from moment to moment, or light physical changes may occur in the stimulus. Distractions or lapses of attention may occur. There may also be errors in giving or recording responses. Experiments should be planned and performed carefully to minimise such errors. To obtain unbiased data when stimuli are presented successively, it is a good idea to present Co before St half of the time and vice versa (remembering to randomise the order).

### ELICITING COMPARATIVE JUDGEMENTS

Where psychophysical techniques elicit the subjective perception of participants in a highly structured way, another method that E/HF specialists use to elicit feelings and opinions from participants in a quantitative manner is by asking them to report their responses directly using a quantitative scale. The reliability, validity and value of such techniques, and indeed the desire and need for quantification, have been a matter of much debate, but these tools remain prevalent in the E/HF toolkit. It is therefore vital that those implementing such methods have a clear understanding of the relationship between the way in which such scales and methods are designed and implemented, and the impact of these design decisions on participant responses.

This section considers a number of commonly applied scaling and quantitative judgement elicitation techniques. A later section in this chapter considers how such scales are implemented in the form of questionnaires or other types of subjective response administration tools.

## RANKING

Sometimes we are particularly interested in the relative order of preference for or attitude towards items. This may be in the context of determining a purchase of workplace equipment or may be to inform theory or future design, for example asking participants to rate their perceptions of the workload placed upon them by different task designs. Therefore, we ask participants to rank items.

The restrictions placed upon the use of ranking techniques are often practical. Reliable ranking is impossible if there are a large number of objects, or if each object takes a long time to consider, for example if you were asking people to rank a list of 20 websites of how easy they were to navigate through. The maximum number of items that participants can normally rank effectively is around nine. Ideally, the objects should all be viewed or experienced at the same time or in very quick succession, and it can be a lot easier to rank things if people either have the physical objects which they can move to visualise their rank position, or some sort of cue cards that can be arranged in order. Many online questionnaire delivery tools also enable items to be sorted into order as part of a questionnaire response.

Ranking does not tend to be very good for comparing very similar objects, for detecting small differences. The data we gather from ranking only give us an idea about the order in which each score occurs (i.e. ordinal data\*). Therefore, we cannot infer anything about the size of the intervals between ranked items for an individual's set of ranked data, but we can convert mean rank scores to interval data by assuming data are normally distributed and then applying a Z-score transformation. Table 4.6 shows an example of this calculation for a hypothetical data set, obtained from participants ranking a set of six control devices for use within a control room setting. First, the rank order of preference for each control device given by each participant is listed and mean rankings for each device calculated. At this stage, the data are ordinal and convey only the averaged preference order for the participant group. Lower scores indicate that more of the participants placed this device at, or near, the top of their preference order, and so it can be seen that preference order for devices in this data set is D, C, F, A, E, B. To convert this to an interval scale, we calculate mean choice (MC), *p* (probability) values, convert the *p* values to *z* scores and then represent these scores on an interval scale from 1 to 100 to show the relative distance between ratings and thus provide a more accessible way of interpreting the relative distances between different items.

The advantage of using a ranking approach is primarily in the form of the ranked list that is obtained at the end of the analysis, which can clearly support decision making, especially in the context of workplace design. In addition, as long as the number of items to be ranked is relatively small, the method can be reasonably easy to apply. However, it is important to ensure that the principle of unidimensionality or homogeneity (that all items are in fact of the same type or dimension) is true, and thus there is some intuitive sense to ranking the items of interest. Even if this is the case, ranking can take quite a long time to implement if there are a large number of items to consider, and even then, people may find it hard to distinguish between some items. Ranking is essentially a 'forced choice' method, and as with the decision about whether to include a neutral option in questionnaire items (see later in this chapter), the E/HF practitioner should think carefully about whether it is in fact reasonable to elicit a forced choice from a participant before applying the ranking method.

---

\* Interval data are that where the distance between points on the scale is equal; ratio data also have this property, but in addition, the scale has a point of 'absolute zero' (easily understood as the point of zero 'meaning nothing' – that is, 0 g represents a quantity that has zero mass). The other two types of data, which do not allow the application of parametric statistics, are nominal (categorical data, where there is no meaningful linear pattern between consecutive points – an example might be a list of different chair styles) and ordinal, where it can be assumed that one point is higher or lower than the one next to it on the scale, but not that the distances between consecutive points are equal. Most statistics text books define these data types in detail.

**TABLE 4.6**  
**Hypothetical Data Set Used as Basis for Ranking Calculations**

| Participant                                                             | Control Device |       |       |      |       |      |
|-------------------------------------------------------------------------|----------------|-------|-------|------|-------|------|
|                                                                         | A              | B     | C     | D    | E     | F    |
| 1                                                                       | 4              | 6     | 3     | 2    | 5     | 1    |
| 2                                                                       | 6              | 5     | 2     | 1    | 4     | 3    |
| 3                                                                       | 5              | 6     | 2     | 1    | 4     | 3    |
| 4                                                                       | 3              | 1     | 2     | 6    | 5     | 4    |
| 5                                                                       | 5              | 6     | 1     | 2    | 4     | 3    |
| 6                                                                       | 3              | 6     | 2     | 1    | 5     | 4    |
| 7                                                                       | 6              | 4     | 3     | 2    | 5     | 1    |
| 8                                                                       | 4              | 5     | 2     | 1    | 6     | 3    |
| Calculate mean ranking (MR)                                             | 4.5            | 4.875 | 2.125 | 2    | 4.75  | 2.75 |
| Calculate mean choice (MC = n-MR)<br>(where n = number of participants) | 3.5            | 3.125 | 5.875 | 6    | 3.25  | 5.25 |
| Calculate p values ( $p = MC/(n-1)$ )                                   | 0.5            | 0.45  | 0.84  | 0.86 | 0.46  | 0.75 |
| Convert p value to z                                                    | 0              | -0.13 | 0.99  | 1.08 | -0.10 | 0.67 |
| Make highest z score = 100, lowest = 0                                  | 11             | 0     | 93    | 100  | 2     | 66   |
| Display ranks proportionally along scale                                |                |       |       |      |       |      |
|                                                                         |                |       |       |      |       |      |

## PAIRED COMPARISONS

The method of paired comparisons overcomes most disadvantages of simple ranking provided that the number of objects the participants have to rank is relatively small (no more than 8–10). All of the objects or items to be considered by the participants are presented in successive pairs, and the participant is asked which is better or larger or preferred (depending on the research question). All possible paired comparisons are used. For example, if there are four objects which are named A, B, C and D, the pairs presented to the participants will be AB, AC, AD, BC, BD, CD. This covers all possible combinations of the items. The two objects of each pair should be presented simultaneously, if at all possible, although sometimes they may have to be presented successively, for example if participants are being asked to rate chairs as they would be unable to sit in two at the same time! In order to avoid bias in the judgements, the following rules for presentation of the objects must be observed:

- Each object should be presented first (or on the left if presented simultaneously) as often as second (or on the right), for example present objects in the order A then B as often as you present them in the order B then A.
- Position or order of presentation of each object should be alternated, for example present object A then B for one participant then object B then A for the next participant.
- No object should be presented in successive pairs, for example present objects A and B then C and D for one participant, but for the next participant, present objects A and C then B and D.

The number of possible pairs of objects is  $n(n-1)/2$ , where  $n$  is the number of objects you want the participants to consider.

In a similar manner to ranking data, calculations can be performed to represent the data obtained along an interval scale; however, rather than calculating mean rankings, a preference matrix is produced, and this is used to identify the mean probability of selecting each item from which the normalised scale values are calculated.

Figure 4.3 shows a worked example of paired comparisons analysis, using hypothetical data elicited from 12 participants who were asked to compare the comfort of each of 6 chair designs. Participants were asked to state which chair they would prefer to use when working at a computer workstation. As six chairs were considered, the number of pairs presented was

$$N * (N - 1)/2 \text{ where } N = \text{number of items presented}$$

$$\text{That is, } 6 * (6 - 1)/2 = 30/2 = 15$$

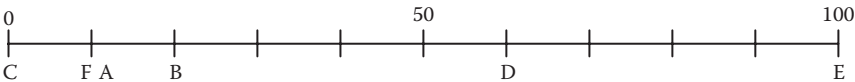
| Preferred choice (figures indicate number of times item in column preferred to item in row, that is, A preferred over B eight times) |       |       |       |      |      |       |
|--------------------------------------------------------------------------------------------------------------------------------------|-------|-------|-------|------|------|-------|
|                                                                                                                                      | A     | B     | C     | D    | E    | F     |
| A                                                                                                                                    | —     | 4     | 1     | 10   | 12   | 3     |
| B                                                                                                                                    | 8     | —     | 4     | 8    | 10   | 6     |
| C                                                                                                                                    | 11    | 8     | —     | 9    | 10   | 7     |
| D                                                                                                                                    | 2     | 4     | 3     | —    | 6    | 2     |
| E                                                                                                                                    | 0     | 2     | 2     | 6    | —    | 1     |
| F                                                                                                                                    | 9     | 6     | 5     | 10   | 11   | —     |
| Probability matrix                                                                                                                   |       |       |       |      |      |       |
|                                                                                                                                      | A     | B     | C     | D    | E    | F     |
| A                                                                                                                                    | 0.50  | 0.33  | 0.08  | 0.83 | 1.00 | 0.25  |
| B                                                                                                                                    | 0.67  | 0.50  | 0.33  | 0.67 | 0.83 | 0.50  |
| C                                                                                                                                    | 0.92  | 0.67  | 0.50  | 0.75 | 0.83 | 0.58  |
| D                                                                                                                                    | 0.17  | 0.33  | 0.25  | 0.50 | 0.50 | 0.17  |
| E                                                                                                                                    | 0.00  | 0.17  | 0.17  | 0.50 | 0.50 | 0.08  |
| F                                                                                                                                    | 0.75  | 0.50  | 0.42  | 0.83 | 0.92 | 0.50  |
| Conversion of p values to z value (note arbitrary value of 4 or -4 awarded to p of 1 or 0)                                           |       |       |       |      |      |       |
|                                                                                                                                      | A     | B     | C     | D    | E    | F     |
| A                                                                                                                                    | 0     | -0.44 | -1.41 | 0.95 | 4    | -0.67 |
| B                                                                                                                                    | 0.44  | 0     | -0.44 | 0.44 | 0.95 | 0     |
| C                                                                                                                                    | 1.41  | 0.44  | 0     | 0.67 | 0.95 | 0.20  |
| D                                                                                                                                    | -0.95 | -0.44 | -0.67 | 0    | 0    | -0.95 |
| E                                                                                                                                    | -4    | -0.95 | -0.95 | 0    | 0    | -1.41 |
| F                                                                                                                                    | 0.67  | 0     | -0.20 | 0.95 | 1.41 | 0     |
| Mean                                                                                                                                 | -0.41 | -0.24 | -0.61 | 0.50 | 1.22 | -0.47 |
| Make highest z score = 100, lowest = 0                                                                                               | 11    | 20    | 0     | 61   | 100  | 8     |
| Display data proportionally along scale                                                                                              |       |       |       |      |      |       |
|                                                  |       |       |       |      |      |       |

FIGURE 4.3 Hypothetical data set for paired comparison calculations.

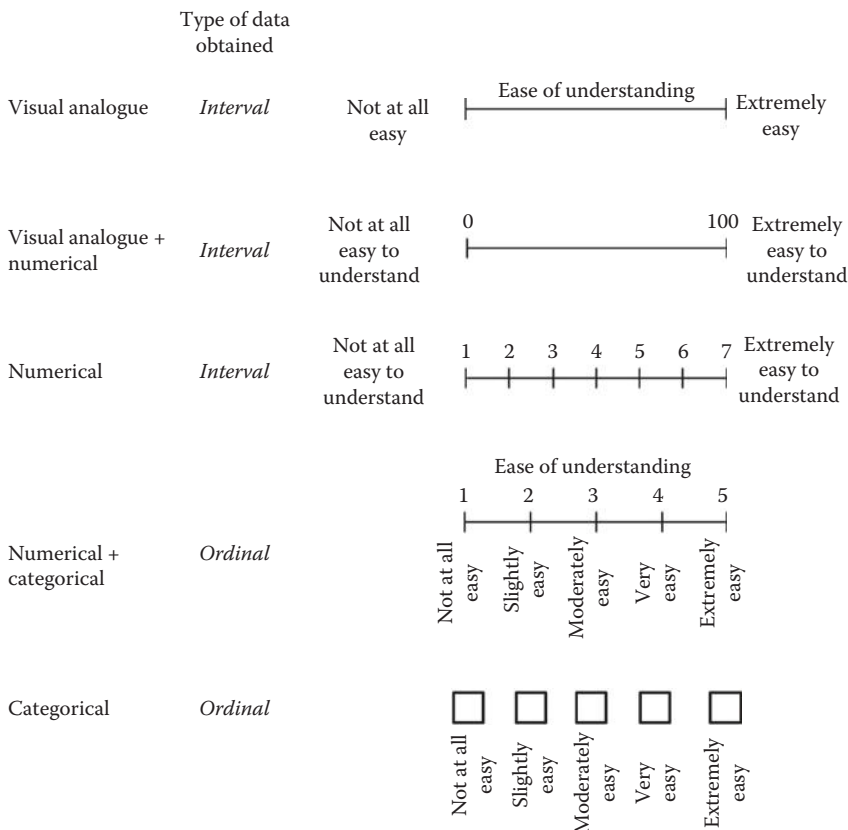
An example order of presentation of pairs is AB, EF, CD, AE, BC, FA, DB, CE, DA, BE, FC, ED, FB, CA, DF. Data from 12 participants are presented.

Advantages of the paired comparisons technique are the reduction of previous question bias by presenting only two items at once, and that participants tend to find it easier to make the simple relative decision of which of two items are preferred. The number of pairs presented can however quite quickly build up, and therefore although from a statistical point of view it is more practical to use this approach to elicit ranks and preferences for large sets of items, from a practical point of view, it can quickly become fatiguing for the participant. In a similar manner to the ranking approach, it is also important that the different items being considered are in fact reasonable to consider along the same feature or characteristic.

## MEASURING OPINIONS AND ATTITUDES: RATING AND SCALING TECHNIQUES

### SIMPLE RATING SCALES

Rating scales are probably the most widely used method of subjective assessment. Each object is rated (or placed) on a given scale, giving it a category or numerical value. The main aim of a rating scale is to allow a respondent to express an opinion or feeling using a quantitative scale. There are a number of ways in which such scales can be designed. Figure 4.4 shows examples of some different ways of presenting five different types of rating scale design that are commonly seen in E/HF investigations. These include specific scale types, such as the visual analogue scale, categorical scale and numerical scale, using the example context of rating 'ease of understanding'



**FIGURE 4.4** Example of different methods of presenting rating scales.



of a product or system. Lee et al. (2010) provide an excellent overview of the different types of scales that might be used to assess subjective perception of thermal sensation.

The choice of rating scale presentation method has implications for ease of understanding by the participant and nature of analysis that is possible once responses have been collected. When considering how the scale appears to the participant, which will contribute to the face validity of the scale (see Chapter 1), a main consideration is whether to include words, either as anchors (end points) or as points on the scale itself.

The inclusion of words can make a scale easier to understand and read, but can also introduce potential concerns about different ways in which the same word might be interpreted by different people.

It therefore should be assumed that, unless specialist techniques such as Thurstone or Likert scaling have been applied (see later in this chapter), the data from scales where words are attached must be treated as ordinal and not interval data, because it cannot be assumed that a respondent will infer equal distances between the different words. We argue, however, that if a scale has only verbal anchor points, and does include numerical points along the scale, the data from such a scale can be treated as interval. In addition, if a scale has been constructed according to certain principles and validated, it can have interval properties so that mean rankings can be calculated and more powerful parametric statistics can be used.

As well as considering the way in which people may interpret distances between points on scales, people may also interpret meanings of words differently. For example, the word 'nausea', which is often used in measures of sickness that might occur in response to using technologies such as simulators or virtual reality (VR), is more commonly used in the United States to mean what someone from the United Kingdom might describe as 'feeling sick'. Although this is a subtle difference, it may be critical when we are interpreting results from participants and in fact, in a shortened version of a scale to measure sickness that was inspired by work originally conducted in the United States, we chose to use the word 'sickness' rather than 'nausea' for use with U.K. participants (Sharpley et al., 2008).

A critical consideration when designing a scale is the analysis which you wish to complete once the data from the scale have been collected. If you wish to apply parametric tests, such as *t* tests or ANOVA (see Chapter 2), then you need to ensure, amongst other assumptions such as homogeneity of variance, independence of sampling and your data being normally distributed, and that data are drawn from a source that is either interval or ratio. In addition to considering the statistical properties of scales, it is also important to consider how easy they are to respond to. A good scale should have high reliability, that is, it should yield the same response from people who have the same feelings or opinions and, if measuring a static characteristic, should yield the same response on repeated administrations. Principles that aid the development of a reliable scale include ensuring that written cues are clear, relevant, precise, objective and unique. A rating scale should be accompanied by a clear statement or question to make it clear what the judgement is about, and whatever type of scale is used, the end points of the scale must always be specified in order to 'anchor' the scale. A particular type of scale that has two opposing end points at the ends of a scale is the 'semantic differential' which can be presented in analogue or numerical format. Anchors should be unambiguous labels which will be interpreted in the same way by all raters. The meanings of anchors (and any other labels used on the scale) are very important. Since there is a tendency for people to avoid the extreme point on any scale, it may be advisable to make the end point descriptions more extreme than required by the tests (particularly for graphic and category scales), for example by using words such as 'extremely'.

There is much discussion about the optimal number of points that should be used within a scale. If cues (subdivisions, either numerical or written) are used on the scale line, the best number of points is normally between 5 and 11. Most scales have 5 or 7 points – on many features, 9 or 11 points are too many since people cannot easily make such fine judgements. Usually, an odd number of items are used, but an even number can be useful to avoid respondents opting for an easy

non-decisive (or neutral) response. A principle that is often useful when deciding whether or not to include a mid- or neutral point is to consider whether or not it is acceptable to allow people to ‘feel’ neutral about the matter being questioned.

However the scale is designed, it is often helpful to follow a rating scale with an open-ended question at which the user can offer opinions or explain the reasons for their judgement. These data cannot be analysed quantitatively, but nevertheless can be very useful – it is often the case that the quantitative data are very useful to explain ‘what’, but the qualitative can explain ‘how’ or ‘why’ (see Chapter 5 for further discussion of qualitative data properties).

## TECHNIQUES FOR GENERATING RATING SCALES

Sometimes, perhaps for reasons of face validity or ease of reporting data, it is desirable to apply a scale that has categorical labels to describe each point, but the investigator also wishes to apply parametric statistics. Therefore, several techniques have been developed to enable the development of scales that have verbal labels for each point, but where it is permissible to apply parametric statistics. This section describes two such methods: the Thurstone scaling technique, which yields a scale that has equal-appearing intervals, and the Likert scaling technique, which does not enable the application of parametric statistics through the assumption of equally appearing intervals, but instead uses a standard agree–disagree response scale and a series of tests of validity and reliability to enable items to be grouped and parametric statistics to be applied to the data that are obtained from combining individual sets of items into groups.\*

### Thurstone’s Method of Equal-Appearing Intervals

The Thurstone method (Thurstone and Chave, 1929) of equal-appearing intervals enables the creation of a single scale, with individual verbal anchors at each point, which can be demonstrated to have equal intervals between each point. The basis of the Thurstone method is the development of a large (around 100) set of items, which vary in the degree of favourableness or unfavourableness about the object, person or system. Methods such as focus groups or interviews with subject matter experts can be used to generate these statements. These statements are then presented to a group of ‘judges’ – ideally as many as 75–100 judges should be recruited, but it is often possible to recruit only fewer. The judges should then sort the statements into 11 piles or categories, which range from one extreme (of favourableness or unfavourableness) to the other. Once this has been done, and after results from any judge who allocated more than 25% of statements to a single file have been removed, the mean and SD of the ratings applied for each statement are calculated. The statements are then allocated to the scale on the basis of selecting those items that are closest to the required intervals (normally a 7- or 9-point scale is generated) and end points, and have the lowest standard deviation (i.e. the highest level of agreement amongst the judges who rated the statements). Pickup et al. (2005) used this approach to develop the unidimensional scale of workload assessment for use within the rail context – the integrated workload scale (see Figure 4.5).

### Likert Scale Development Analysis

A method that is commonly used to generate and apply a set of items to measure a particular construct is the Likert scaling technique. The phrase ‘Likert scale’ is perhaps one of the most commonly misused phrases within E/HF, as it strictly applies only to a scale that is developed with the specific ‘strongly agree–strongly disagree’ response scale. A standard rating scale (e.g. as in the categorical scale example presented earlier in Figure 4.4) should be described simply as a ‘rating scale’. This section describes the steps involved in developing a Likert scale;

---

\* There is in fact much debate about the treatment of Likert scales as interval or ordinal data. The authors of this chapter take the view that as long as appropriate tests for validity and reliability have been applied, it is reasonable to apply parametric statistics to data from Likert scales with caution. See Norman (2010) for a particularly robust dismissal of concerns about treatment of data from Likert scales as interval data.

|                                                                                   |                       |                                                                                                                            |
|-----------------------------------------------------------------------------------|-----------------------|----------------------------------------------------------------------------------------------------------------------------|
|  | Not demanding         | Work is not demanding at all                                                                                               |
|  | Minimal effort        | Minimal effort required to keep on top of situation                                                                        |
|  | Some spare time       | Active with some spare time to complete less essential jobs                                                                |
|  | Moderate effort       | Work demanding but manageable with moderate effort                                                                         |
|  | Moderate pressure     | Moderate pressure, work is manageable                                                                                      |
|  | Very busy             | Very busy but still able to do job                                                                                         |
|  | Extreme effort        | Extreme effort and concentration necessary to ensure everything gets done                                                  |
|  | Struggling to keep up | Very high level of effort and demand, struggling to keep up with everything                                                |
|  | Work too demanding    | Work too demanding – complex or multiple problems to deal with and even with very high levels of effort it is unmanageable |

**FIGURE 4.5** Pickup et al. (2005) integrated workload scale: developed using Thurstone's scaling technique.

these steps can also be applied to sets of items with responses on identical scales, to confirm question set reliability and validity.

The process of generating a Likert scale, as with the Thurstone approach, involves a series of stages and recruitment of 'judges'. A large set of items is generated – it is recommended that when determining the number of items to be generated, the researcher should be confident that they have achieved 'data saturation', that is, they are unable to identify any new items related to the concept under investigation. Item sets should also include both positively and negatively worded items – this avoids previous question bias (where a participant tends to copy patterns of responses from earlier questions) in questionnaire responses. This generation of items may be supported by activities such as focus groups or subject matter expert interviews. An underlying principle of the Likert approach is that either a single psychological construct is being investigated or a coherent group of related constructs are being captured (i.e. unidimensionality). For example, when Nichols (1998) developed a Likert-based questionnaire to measure VR attitudes, she assumed (1) that there was a sense of 'attitude towards VR' that was meaningful to capture, and (2) that this attitude would comprise a set of sub-constructs.

Once the E/HF expert is confident that the item set represents all the possible constructs and elements, these items should be put into the form of a questionnaire. This questionnaire should then be administered to participants, accompanied by the selected length of strongly agree–strongly disagree scale (with numerical values attached to each point for input into analysis spreadsheets), using the anticipated number of factors and initial number of items as a basis for determining the number of responses which should be obtained. Ferguson and Cox (1993) note that the participant-to-items ratio should be between 2:1 and 10:1, the absolute minimum number of participants should be 100, and the relative minimum proportions of items to expected factors and participants to expected factors should be between 2:1 and 6:1. It is important to note that the purpose of this initial administration of the questionnaire is not to measure participant attitudes or feelings about the items, but to generate data to enable statistical analysis; therefore, it is often sensible to aim to administer the questionnaire to a wide range of participant types (as wide as would be reflected by the eventual participants to be investigated). Similar to the approach of Pickup et al. (2005), who allowed respondents to the Thurstone items to note whether any were 'confusing', Nichols (1998) included a '?' option to allow participants to note when items were difficult to understand.

After data have been collected from the initial set of participants, the direction of scoring for each item should be determined. Initially, this can be estimated by observation, and the numerical score for any negatively worded questions reversed. A simple calculation to enable reversal of scorings is to subtract the score from the maximum point +1. In other words, if a scale is scored from 1 to 5, if a participant responds with a score of 2, if that item is determined to be 'negative', it can be converted

to its positive value by subtracting 2 from the maximum scale score (5) plus 1 (i.e.  $6 - 2 = 4$ ). It is good practice to always work on the basis that a high score represents a 'good' or 'positive' value – intuitively, we read high values on graphs or tables as being positive, and it is useful to have this as an underlying and consistent rule throughout all analysis. If there is any doubt about the appropriate direction of an item at this stage, it should be noted as 'neutral' and retain positive direction scoring, until later statistical analysis can confirm whether this designation of direction was correct.

At this point, the E/HF expert should decide how to deal with any items where a high number of participants selected the '?' option or indicated that they were difficult to understand. Normally such items would at this stage be omitted, but if there is any specific reason that the content of the items is important to the topic of interest, it may be appropriate to reword the items at this stage; strictly speaking, the process of obtaining data from participants on the new items should then be repeated, but there is often a point where repeated administrations of the questionnaire become impractical, and it may be appropriate to base any small question refinements on expert judgement and, with caution, retain them for inclusion in later administrations. It may also be appropriate to remove all responses of any participant who has given a high number of '?' responses.

The next activity that should be completed is further analysis of the neutral items. This can be done by obtaining the total score for each participant for all those items where the researcher is confident of the direction of scoring and correlating the individual response to each neutral item with the subtotal (this is essentially a modified version of the item-total coefficient that is used as the basis of Cronbach's alpha reliability analysis as described later). For those items that have a significant positive or negative correlation with the subtotal, this can be used as the basis for determining direction of scoring; if any neutral items remain, they should at this point either be discarded or treated as positive (it is quite likely that there is something imperfect about the wording or relevance of these items, and that they will be omitted in later tests of validity within factor analysis). Nichols (1999) at this stage also conducted analysis of the SD of items and the value of the mean, to examine evidence for any other issues such as central tendency.

Once direction of scoring for all items has been determined, reliability analysis should be conducted. This normally consists of two stages, item analysis (where individual items are correlated with the scale total minus that item) and an overall statistic known as Cronbach's alpha. Nunally (1978) suggests that in the case where it is suspected that a number of factors may be present within a scale, item analysis (correlation of individual items with the total score or the total score minus that item) can be used to identify exogenous variables to be omitted and to show the presence of a general factor, which is an artefact rather than a result of a complete factor analysis. Cronbach's alpha is an overall correlation coefficient, which indicates a high level of reliability and internal consistency in a test. Kline (1994) recommends that a reliability coefficient of at least 0.80 is desirable, but Lowenthal (1996) warns that such a high reliability coefficient may result from a repetitious questionnaire and that it is also important to ensure that a test has high validity, has a sound theoretical basis and as few items as possible. When a test of reliability is performed, a corrected item-total correlation coefficient is obtained for each item in the questionnaire. Those items with very low correlations should be removed from the test. Lowenthal (1996) recommends that items with a correlation of below 0.15 should definitely be removed, unless no higher correlations are available. The removal of these items contributes to the construct validity of the scale (see Chapter 1).

The final stage of development of the Likert scale is to conduct a factor analysis. There are several excellent texts that discuss factor analysis in much more detail than is possible here (see Kline, 1994).

Once the data have been tested to ensure they meet univariate normality, a standard method of factor analysis, such as principal components analysis (often applied with a rotation such as orthogonal varimax) is applied. The number of factors to be extracted can be determined by a number of methods, including selection of factors with an 'eigen value' (a value that describes the coherence of items within the identified factor) of greater than 1 or by using a 'scree plot', which is a line graph of eigen values of factors in descending order and asks the researcher to note where there is a change in gradient of the line.

An example of the data that then emerges from such an analysis is shown in Table 4.7.

**TABLE 4.7**  
**Factor Loadings of Questionnaire Items<sup>a</sup>**

| Question            |                               | Factor      |             |             |             |             |             |             |             |
|---------------------|-------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                     |                               | 1           | 2           | 3           | 4           | 5           | 6           | 7           | 8           |
| 3                   | Health anxiety – 1            | <i>0.44</i> | 0.03        | <i>0.46</i> | 0.01        | 0.34        | 0.23        | –0.09       | 0.04        |
| 20                  | Health anxiety – 2            | <i>0.76</i> | 0.04        | 0.14        | 0.09        | 0.12        | 0.08        | 0.12        | –0.03       |
| 29                  | Health anxiety – 3            | <i>0.83</i> | 0.16        | 0.06        | 0.09        | 0.05        | 0.02        | 0.21        | 0.02        |
| 30                  | Health anxiety – 4            | <i>0.63</i> | 0.04        | 0.11        | –0.03       | 0.09        | 0.17        | 0.36        | 0.12        |
| 31                  | Health anxiety – 5            | <i>0.48</i> | 0.01        | –0.05       | 0.14        | 0.17        | <i>0.47</i> | 0.14        | –0.03       |
| 41                  | Health anxiety – 6            | <i>0.72</i> | 0.21        | –0.03       | –0.06       | 0.03        | 0.36        | –0.08       | 0.12        |
| 1                   | VR desirability – 1           | 0.15        | <i>0.70</i> | 0.14        | 0.04        | 0.36        | –0.05       | 0.02        | –0.01       |
| 5                   | VR desirability – 2           | –0.08       | <i>0.66</i> | 0.18        | 0.26        | –0.04       | –0.29       | 0.01        | 0.08        |
| 9                   | VR desirability – 3           | 0.10        | <i>0.77</i> | 0.16        | 0.14        | –0.02       | –0.01       | –0.01       | 0.14        |
| 11                  | VR desirability – 4           | 0.29        | <i>0.67</i> | 0.14        | 0.18        | 0.34        | 0.06        | –0.05       | –0.02       |
| 21                  | VR desirability – 5           | –0.13       | <i>0.52</i> | –0.01       | –0.29       | 0.02        | 0.17        | 0.21        | 0.06        |
| 45                  | VR desirability – 6           | 0.07        | <i>0.48</i> | –0.07       | 0.28        | 0.21        | 0.21        | –0.13       | 0.32        |
| 3                   | VR awareness – 1              | <i>0.44</i> | 0.03        | <i>0.46</i> | 0.01        | 0.34        | 0.23        | –0.09       | 0.04        |
| 12                  | VR awareness – 2              | 0.27        | –0.04       | <i>0.57</i> | 0.05        | 0.15        | 0.35        | –0.22       | 0.18        |
| 13                  | VR awareness – 3              | 0.31        | –0.06       | <i>0.44</i> | 0.21        | 0.27        | 0.11        | –0.12       | 0.39        |
| 18                  | VR awareness – 4              | 0.26        | 0.25        | <i>0.49</i> | 0.25        | 0.23        | –0.10       | 0.08        | –0.08       |
| 22                  | VR awareness – 5              | 0.02        | 0.11        | <i>0.70</i> | 0.09        | –0.06       | 0.11        | –0.06       | –0.03       |
| 33                  | VR awareness – 6              | –0.13       | 0.28        | <i>0.55</i> | –0.12       | 0.17        | –0.01       | 0.16        | 0.23        |
| 36                  | VR awareness – 7              | –0.05       | 0.29        | <i>0.58</i> | 0.12        | –0.01       | 0.11        | 0.17        | –0.08       |
| 40                  | VR awareness – 8              | 0.38        | –0.06       | <i>0.53</i> | 0.22        | 0.34        | 0.07        | –0.11       | 0.11        |
| 6                   | VR usefulness – 1             | –0.03       | 0.37        | 0.10        | <i>0.41</i> | 0.35        | –0.01       | 0.11        | 0.28        |
| 7                   | VR usefulness – 2             | 0.12        | 0.09        | –0.16       | <i>0.66</i> | 0.12        | 0.32        | –0.06       | –0.15       |
| 32                  | VR usefulness – 3             | –0.08       | 0.16        | 0.13        | <i>0.70</i> | 0.14        | 0.14        | –0.18       | 0.30        |
| 37                  | VR usefulness – 4             | 0.09        | 0.11        | 0.19        | <i>0.82</i> | –0.03       | –0.16       | 0.21        | –0.01       |
| 38                  | VR usefulness – 5             | 0.09        | 0.15        | 0.26        | <i>0.80</i> | 0.03        | –0.17       | 0.12        | 0.10        |
| 14                  | Enthusiasm for VR – 1         | 0.21        | 0.32        | 0.01        | 0.10        | <i>0.75</i> | 0.11        | –0.03       | 0.04        |
| 23                  | Enthusiasm for VR – 2         | 0.09        | 0.08        | 0.22        | 0.03        | <i>0.62</i> | –0.18       | 0.20        | 0.26        |
| 27                  | Enthusiasm for VR – 3         | 0.00        | 0.19        | 0.33        | 0.11        | <i>0.59</i> | 0.17        | 0.13        | –0.15       |
| 8                   | Safety of VR – 1              | 0.23        | 0.15        | –0.08       | –0.15       | 0.03        | <i>0.65</i> | 0.23        | 0.10        |
| 16                  | Safety of VR – 2              | 0.11        | –0.04       | 0.24        | –0.01       | –0.04       | <i>0.69</i> | 0.11        | 0.09        |
| 31                  | Safety of VR – 3              | <i>0.48</i> | 0.01        | –0.05       | 0.14        | 0.17        | <i>0.47</i> | 0.14        | –0.03       |
| 43                  | Safety of VR – 4              | 0.16        | –0.02       | 0.30        | 0.06        | 0.05        | <i>0.61</i> | 0.29        | –0.10       |
| 2                   | Immersion and headset use – 1 | 0.21        | 0.16        | –0.27       | 0.07        | 0.15        | 0.03        | <i>0.53</i> | 0.32        |
| 4                   | Immersion and headset use – 2 | 0.15        | 0.02        | –0.05       | 0.01        | 0.16        | 0.11        | <i>0.72</i> | –0.08       |
| 35                  | Immersion and headset use – 3 | 0.28        | –0.02       | 0.26        | –0.10       | –0.06       | 0.23        | <i>0.58</i> | 0.18        |
| 25                  | Advantages of VR – 1          | 0.20        | 0.23        | 0.14        | –0.03       | –0.15       | –0.07       | –0.06       | <i>0.71</i> |
| 34                  | Advantages of VR – 2          | –0.06       | 0.05        | –0.08       | 0.18        | 0.26        | 0.10        | 0.15        | <i>0.64</i> |
| Eigen values        |                               | 8.08        | 3.61        | 2.23        | 2.07        | 1.85        | 1.63        | 1.50        | 1.25        |
| Cumulative variance |                               | 21.3        | 30.8        | 36.6        | 42.1        | 46.9        | 51.2        | 55.2        | 58.5        |

Source: Nichols, S., Virtual Reality Induced Symptoms and Effects: Theoretical and methodological issues, PhD thesis, University of Nottingham, Nottingham, U.K., 1999.

<sup>a</sup> Figures in italics denote factors onto which each item loads (values over 0.4).

The investigator then examines the loadings of items onto different factors. The correlation coefficient usually adopted for the selection of appropriate variables loading onto a factor is 0.30. However, this is often modified in some way to account for the possibility of individual variables loading onto more than one factor (cross-loadings). Igarria et al. (1994) require that a given item should load 0.50 or higher on a specific factor and have a loading no higher than 0.25 on other factors. However, Ferguson and Cox (1993) recommend a loading of 0.4 in an endeavour to reach high factor saturation (high mean factor loadings for a factor). A criterion that an item should have a loading of  $>0.40$  is therefore recommended.

Adopting this higher cut-off point also eliminates a number of problems associated with cross-loadings. For example, in the Nichols (1999) data, there are only two variables for which cross-loading occurs: questions 3 and 31. Ferguson and Cox (1993) suggest that if indeed the factors are required to be clear and distinct, then if the difference between the two loadings is small ( $<0.2$ ), the variable should be removed. However, if the difference is  $>0.2$ , the variable can remain and be assumed to load onto the factor with the highest loading. On the other hand, the occurrence of cross-loading may be indicative of conceptual overlap, where the two variables do indeed contribute to more than one factor.

The final stage of Likert scale development is to name the factors. This is achieved by presenting the items in groups to independent judges and asking the judges to come up with names to describe the item groups; this can work particularly well as a group exercise, where E/HF experts act as the judges and collaboratively develop names that they are comfortable represent the item groupings.

## QUESTIONNAIRES

Once the individual scales and questions that will be asked within a study have been identified, they need to be delivered to the participant in an appropriate form. This may involve the production of a questionnaire, but may also use techniques such as interviews and diary studies, or even involve tools that the experimenter themselves completes, such as structured observations. This section describes these different approaches and considers their advantages and pitfalls when applied within E/HF studies.

Questionnaires are probably the most frequently used method within E/HF, used to collect participant experiences, requirements and opinions. There are many excellent texts and resources on questionnaire design (e.g. Oppenheim, 1992), but key principles are outlined here.

Increasingly, questionnaires are administered through electronic tools. Examples of these that are currently commonly used in the United Kingdom include Survey Monkey, Qualtrics and the Bristol Online survey.\* The use of online methods to distribute questionnaires has considerable advantage in removing the need for transfer of data from the paper-based form to an analysis tool or spreadsheet, enables wide geographical access to participants and can provide respondents with considerable flexibility with how and when they respond to the questionnaire (e.g. by enabling the questionnaire to be viewed on a tablet or smartphone device). However, they should not be seen as the panacea – if a questionnaire is distributed online, then it is not easily possible to confirm who has responded, there may be bias in sampling towards participants who feel comfortable and confident responding to a questionnaire online and, critically, there are differences in how online questionnaire tool providers store and use data that are administered through their distribution tools. Therefore, any E/HF practitioner who is using such tools should be completely aware of how their tools are distributed and accessed, and how the data from their questionnaire are stored in the

---

\* Currently to be found at <http://www.surveymonkey.com>, <http://www.qualtrics.com>, <http://www.survey.bris.ac.uk>. Accessed November 2014.



short and long terms. Such issues associated with data storage also apply to the storage of any E/HF data online using cloud or distributed storage systems.

## **INSTRUCTIONS AND BACKGROUND INFORMATION**

All questionnaires will include an introduction explaining the purpose of the questionnaire. Such sections should include contact details for the questionnaire owner and give clear and accurate information about how data from questionnaires are going to be stored, analysed and shared. A common mistake at this point is to promise something which will not be delivered – for example will the data be accessed only by the investigator themselves or will their colleagues or boss also see the data? A common confusion at this point is to state that all data will be ‘anonymous’. True anonymity means that the data will not under any circumstances enable the original respondent to be identified; recent release of public data and numerous public news stories have demonstrated the phenomenon of ‘jigsaw identification’ where relatively few data points can be used to uniquely identify an individual. As an example, if a questionnaire asks participants to state which age category they are in and also which work group in an organisation that they belong to, it may be possible that these data are sufficient to identify individuals. Even though the investigator may not use the information in this way, if participants realise (or fear) that they can be identified, they may prefer not to complete the questionnaire. Therefore, it is often the case that a process must be put in place to ensure ‘confidentiality’ – being very clear about where data are going to be released and in what form. This can be particularly important when collecting data from within a workplace context, for example if we are interested in job satisfaction and stress in a workplace, and want to capture information about causative influences, it is important to reassure respondents that they can respond honestly and their responses will not be used negatively towards them by colleagues or superiors. This may demand anonymity, but may also demand that some data are not directly shared in the reporting of the results.

## **QUESTIONNAIRE FORMAT**

If a questionnaire appears clear, attractive and professional, respondents will be more keen to respond – they will feel valued, will feel that effort has been made in producing the questionnaire and will feel more confident that they will be able to understand the questions being asked. Online tools as mentioned earlier can be very helpful in delivering a professionally presented questionnaire (although of course a polished appearance cannot compensate for unclear question wording!). A rule of thumb is that a questionnaire should be as short as possible – for every single question, the investigator should think carefully about whether that question is actually needed. Commonly included, possibly, unnecessary questions include those asking for information about educational background, age and job role. Whilst this can of course in some cases be critical information to aid interpretation of results, they are often included in questionnaires by default. The inclusion of educational background and job role can be surprisingly difficult to collect and analyse – for example the words ‘school’ and ‘college’ have different meanings in the United States and the United Kingdom and, if the questionnaire is to be responded to by an international sample, may mean that responses are difficult to interpret. Similarly, unless specifically applying a tool within a single organisation, capture of information about job role can quickly become ambiguous – for example in the United Kingdom, the job of a librarian can be known as an information officer, information specialist or information scientist. Questions such as age can not only sometimes be unnecessary, but can also be off-putting to respondents, who may not wish to disclose information that they perceive as personal (see next section for further discussion of the inclusion of personal questions in questionnaires).



Finally, the ‘tone’ of the questionnaire should be carefully considered and communicated – if this is a questionnaire for completion by those who may have limited language ability (e.g. children, respondents with disabilities or respondents reading in their second language), this should be carefully considered and simple language used throughout (an online readability calculator can be used to check readability using scores such as a Flesch score [Flesch, 1948]).

## QUESTION CONTENT

Some advice on question content has already been provided in specific parts of this chapter that described rating scales, but there are some general considerations that should also be addressed. First, the questionnaire designer should carefully ensure that all questions can and will be answered by respondents. Elements that can be off-putting are questions that ask about information that the respondents might be embarrassed to disclose (e.g. personal information such as weight) or which are seen as an invasion of privacy (e.g. asking about income). If such questions are deemed necessary, it is particularly important to consider the issues of data confidentiality and anonymity as described in the ethics section. If personal questions are necessary, they should be placed at the end of the questionnaire, and, if possible, be optional.

In addition to the quantitative question types described elsewhere in this chapter, the reader is encouraged to consider the inclusion of open-ended questions within their questionnaires, particularly if it is important to understand ‘how’ or ‘why’ the participant has specific attitudes or opinions. In our experience, the data from such questions can be particularly valuable in explaining the ‘how’ and ‘why’ behind quantitative data obtained.

## QUESTION WORDING

In addition to taking into account some of the matters discussed earlier with respect to rating scale designs, there are specific considerations that can be made regarding the wording of individual question items. Questions should be specific, simple, avoid ambiguities and, if technical jargon is included, should define the meaning of words where necessary. It is dangerous to assume however that a respondent will read any explanatory text, so the wording of questions must be very carefully tested. Participants can demonstrate previous question bias (where they alter their responses to agree with those previously given) and may also be reluctant to display responses that use the end points of the question scales; this may be a motivation to include a 7- or 9-point scale, rather than a shorter scale, or even to remove the mid-point of the scale completely. Participants may also exhibit ‘prestige bias’ (Sinclair, 2005) where they are aware of the perception of the response they give and so, for example, may adjust the way in which they report views or attitudes on socially desirable or contentious issues, reporting perhaps what they aspire to think or do, rather than the true state of affairs. It is important to note that in some industrial contexts, it is in fact important to use technical jargon and terms – this helps to emphasise the specific application of the questionnaire to that particular participant group and may encourage a higher response rate. If such terms are to be used, it is extremely important that the questionnaire is reviewed not only by E/HF experts but also by those familiar with the domain in which the tool is to be applied. Finally, an obvious but occasionally overlooked principle of question design is to avoid leading questions. An example of a leading question might be ‘How do you think the thermal environment in your office could be improved?’ – this wording perhaps implies that there is room for improvement, and the thermal environment is not currently optimal. A better phrasing of this question might be ‘Do you think that the thermal environment in your office needs to be changed?’ This does not lead the respondent towards the suggestion that the environment should be improved and so is more objective. However, single-word responses (such as ‘yes’) will not help the E/HF researcher to understand what should be changed. For this type of question, it is useful to include an open-ended follow-on question for

the respondent to provide additional comments. Questions such as ‘Do you think your pay level, which is informed by the recent government pay review, is satisfactory?’ (this question could elicit an opinion about level of pay OR attitude towards the pay review process) which have double content should also be avoided.

### **PRACTICAL ISSUES**

There are some practical issues to consider in questionnaire administration. Response rates can be as low as 10%–20%, so if an absolute number of responses is needed, the distribution strategy should be carefully planned. The timing of the process should also be planned, with sufficient time built in for chasing up after low response rates, and inputting, cleansing and analysing data. It can be advantageous to arrange administration of the questionnaire as part of a normal work process to increase response rates – for example Ryan et al. (2009) administered a questionnaire within the U.K. rail community as part of routine safety briefings and thus received an impressive sample size of almost 4000 with a response rate of 83%. Finally, we do not believe it is possible to over-pilot a questionnaire. Most successful questionnaires are developed by teams, because it can take many perspectives to consider all ways in which prompts can be misinterpreted or confusion can be reached. Therefore, building in time for piloting and revising questionnaire-based data collection tools is paramount.

### **GENERAL CONSIDERATIONS IN COLLECTING AND OBSERVING PARTICIPANT RESPONSES**

This chapter has outlined some of the key considerations when designing and applying methods for collecting and observing participant responses. In addition to these specific considerations associated with particular methods, a number of common issues apply to all of these methods.

Increasingly, methods are available to us that allow us to conduct a study without the participant being explicitly aware; for example many workplaces automatically log interactions and, for security and incident recording purposes, are equipped with CCTV, and our interaction with online and cloud-based systems is normally stored, contributing to our ‘lifelong contextual footprint’. But the use of these data to analyse work presents ethical challenges. We normally work according to the principle of ‘informed consent’ – a participant, whether taking part in a research study or being studied as part of a workplace intervention, should be aware of what data are being collected, how they are being stored, what interpretations will be made from the data and how the data will be reported. This principle needs to be retained, even as the collection of workplace and task performance has the opportunity to become more implicit.

We also frequently now collect both quantitative and qualitative data when we are not necessarily aware of what we will find from our data once we begin our analysis. We have moved beyond a solely ‘hypothesis-driven’ approach, but therefore need to be prepared at the outset of a study for what we will do if we find something that we have not perhaps expected. If, for example, we are asked to review a particular worksite that has had notably good performance, to help a company understand how to improve performance across the workforce, but we find that in this ‘high-performing’ worksite there are in fact a number of safety violations occurring, how should we report and manage this? It is important when preparing a study to ‘expect the unexpected’ and work with clients or colleagues to anticipate any information that might be gleaned and plan in advance how this will be managed.

Technology provides increased opportunities for data collection and storage, but this means that our methods of storing and disposing of data should be robust. We are sure that there are many E/HF practitioners around the world who still keenly store data from studies conducted 20+ years

ago (and some of that data may well be in forms such as videotapes that are unlikely to be in a condition that would allow them to be accessed anyway!). We should ensure that we have clear policies and infrastructure to robustly store, access and dispose of data according to both ethics requirements and data protection guidelines that may well be local to a particular country. E/HF as a discipline has not taken the approach of other disciplines, such as climate science, of ‘data sharing’ workspaces, and such an approach could well present complex issues of true data anonymity (remembering how easy it can be to identify an individual from relatively few data points), but it would be naïve to ignore discussion of making not only our publication of work ‘open access’ but also the data collected to underpin our findings.

The quantitative vs. qualitative debate is still relevant (see Chapter 1 for further discussion), but increasingly, E/HF analysis takes a mixed-methods approach, where different tools are used to capture different aspects of work or tasks. Taking into account the context of study can help to provide a better understanding of participant responses (Bisantz and Drury, 2005), and identification of user opinion using different methods (e.g. survey methods and psychophysics studies) can help to identify what factors contribute to user perceptions of products or workplace issues (Liu, 2003). When designing a data collection exercise, the practitioner should be aware of the expectations of those involved with the project for the quantification of data. Our experience tells us that the form of presentation of data, perhaps using quantitative graphs, or digital human models (see Chapter 8), can affect the perceived confidence that people might have in that data. But quotes and scenarios drawn from qualitative sources can also be very powerful communication tools, and the mode of reporting data should always be carefully selected based on the needs and knowledge of the recipients as well as the quality and form of the original data.

Many of our methods and analytical tools tend towards the quest to find a ‘single solution’ – indeed in many work contexts, it is practical to propose only one new technology system or one new workplace design. However, it is important that any E/HF methods capture variation as well as consensus. Some traditional methods of reporting statistics, such as the use of the mean, tend towards consensus reporting, and it is critical that we also report and understand any variation or anomalies in our data. At the very least, this means using metrics such as standard deviation in reporting data, but it is also important to understand *why* an individual or group of individuals appear to have a different view from the ‘norm’. This is consistent with our systems perspective to E/HF analysis – the reason a group might be reporting lower levels of physical symptoms in response to work might, of course, not only be due to the physical layout or demand of their work, but also because they are influenced by the organisational climate or local team culture (see Chapters 29 and 32). Some empathic methods used in inclusive design (see Chapter 11) specifically address this challenge by focusing on designing for specific users with characteristics that may present a particular design challenge.

## CONCLUSIONS

As with other chapters in this volume, this chapter does not aim to provide a ‘how-to’ manual for collecting and observing participant responses; instead, we hope to arm the E/HF practitioner with the necessary knowledge to be able to select the correct method for the correct situation. In most cases, methods are not applied in isolation, so the interaction effects of applying multiple methods also need to be considered. It is also important that the method of data collection is considered at the same time of as the requirements for data reporting – sometimes we can be tempted to apply complex analysis and fancy graphical presentation to our data to communicate it, but we should remember that our lowest common denominator is the method we used for collecting that data in the first place, and it is only by applying rigour and thought to how we collect data from participants in the most appropriate manner that we can ensure that our findings and inferences are as useful and meaningful as possible.

## APPENDIX: APPLICATION FOR APPROVAL OF RESEARCH STUDY INVOLVING HUMAN PARTICIPANTS USED IN THE UNIVERSITY OF NOTTINGHAM FACULTY OF ENGINEERING

### ALL applicants must provide the following information:

The applicant must be the person who will conduct the investigations; each application must be made by one applicant:

- Usually the student in the case of taught or research courses
- Usually the researcher (the member of university research or academic staff) who will conduct the study in the case of funded research projects
- Usually the principal investigator in the case of applications for ethics approval in advance of submission of a research proposal

**If the applicant is an Undergraduate or Postgraduate taught or research student, please complete the following information. The application must be approved by a supervisor.**

|                  |  |                |  |
|------------------|--|----------------|--|
| Name of Student: |  | Student No:    |  |
| Course of Study: |  | Email Address: |  |
| Supervisor:      |  |                |  |

**If the applicant is a member of university research or academic staff, please complete the following information. For research staff, the application must be approved by the principal investigator.**

|                |  |                                           |  |
|----------------|--|-------------------------------------------|--|
| Name:          |  | Principal Investigator<br>(Budget Holder) |  |
| Email Address: |  | PI Signature: .....                       |  |

**Title of investigation:** .....

Planned date for study to begin ..... Duration of Study .....

**Please state whether this application is**

☐ New    ☐ Revised    ☐ A renewal    ☐ For a continuation study

### SELECTION OF REVIEW PROCESS

Please indicate whether the application is required to go forward to the ethics committee for formal review or, in the case of projects completed by *taught undergraduate and postgraduate students only*, whether the application can be approved under expedited review.

☐ Formal review, application will be submitted to ethics committee    ☐ Expedited review, application is approved by supervisor\*

\*This option can be selected only if the supervisor is a member of the Faculty Ethics committee.

## ETHICAL ISSUES CHECKLIST

The purpose of this checklist is to facilitate the review process and to identify any ethical issues that may concern the committee. It is meant to be an aid to both the researcher and the committee. Listed in the following text are areas which require some justification and attention on your part in specifying your study protocol. Please answer each question honestly, giving full details where required. Answering YES to any of the questions will not necessarily lead to a negative response to your application, but it will draw issues to your attention and give the reviewers the opportunity to ensure appropriate steps are being taken. In expedited review, supervisors should ensure that for any questions where the answer YES has been given, appropriate measures have been taken to maintain ethical compliance.

Applicant's full name .....

**You must complete ALL of this section before submitting your application.**

- 1 Who is the population to be studied?  
.....
- 2 Please give details of how the participants will be identified, approached and recruited. (Include any relationship between the investigator and participants, e.g. instructor–student.)  
.....  
.....
- 3 Will the population studied include any vulnerable members of the public?      **YES**      **NO**  
Note: for the purpose of ethics approval, this includes participants who are under 18, people who are disabled or in poor health, and also those who are non-English speakers and may not be able to understand the consent forms. (If YES, please give further details.)  
☐      ☐  
.....
- 4 Will it be possible to associate specific information in your records with specific participants on the basis of name, position or other identifying information contained in your records?      **YES**      **NO**  
☐      ☐
- 5 What steps have you taken to ensure confidentiality of personal information and anonymity of data both during the study and in the release of its findings?  
.....  
.....
- 6 Describe what data will be stored, where, for what period of time, the measures that will be put in place to ensure security of the data, who will have access to the data, and the method and timing of disposal of the data.  
.....  
.....  
.....  
.....  
.....  
.....

|       |                                                                                                                                                                                                                       |                                 |                                |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|--------------------------------|
| 7     | Will persons participating in the study be subjected to physical or psychological discomfort, pain or aversive stimuli which are more than expected from everyday life? (If YES, please give further details)         | YES<br><input type="checkbox"/> | NO<br><input type="checkbox"/> |
| ..... |                                                                                                                                                                                                                       |                                 |                                |
| 8     | Will participants engage in strenuous or unaccustomed physical activity? (If YES, please give further details)                                                                                                        | YES<br><input type="checkbox"/> | NO<br><input type="checkbox"/> |
| ..... |                                                                                                                                                                                                                       |                                 |                                |
| 9     | Will the investigation use procedures designed to induce participants to act contrary to their wishes? (If YES, please give further details.)                                                                         | YES<br><input type="checkbox"/> | NO<br><input type="checkbox"/> |
| ..... |                                                                                                                                                                                                                       |                                 |                                |
| 10    | Will the investigation use procedures designed to induce embarrassment, humiliation, lowered self-esteem, guilt, conflict, anger, discouragement or other emotional reactions? (If YES, please give further details.) | YES<br><input type="checkbox"/> | NO<br><input type="checkbox"/> |
| ..... |                                                                                                                                                                                                                       |                                 |                                |
| 11    | Will participants be induced to disclose information of an intimate or otherwise sensitive nature? (If YES, please give further details.)                                                                             | YES<br><input type="checkbox"/> | NO<br><input type="checkbox"/> |
| ..... |                                                                                                                                                                                                                       |                                 |                                |
| 12    | Will participants be deceived or actively misled in any manner? (If YES, please give further details.)                                                                                                                | YES<br><input type="checkbox"/> | NO<br><input type="checkbox"/> |
| ..... |                                                                                                                                                                                                                       |                                 |                                |
| 13    | Will information be withheld from participants that they might reasonably expect to receive? (If YES, please give further details.)                                                                                   | YES<br><input type="checkbox"/> | NO<br><input type="checkbox"/> |
| ..... |                                                                                                                                                                                                                       |                                 |                                |
| 14    | Will the research involve potentially sensitive topics? (If YES, please give further details.)                                                                                                                        | YES<br><input type="checkbox"/> | NO<br><input type="checkbox"/> |
| ..... |                                                                                                                                                                                                                       |                                 |                                |

If you require space for additional information, please add it here and identify the question to which it refers:

### Checklist of information to include with your application

Please tick the boxes given to confirm that you have included the following information with your submission. Failure to include the required information may result in your ethics application and approval for start of your research to be delayed.

- ☐ A brief description of the study design
- Number and type of participants
  - Number and duration of activities participants will be involved in
  - Equipment and procedures to be applied
  - Information about how participants will be recruited
  - Whether participants will be paid (state how this will be done)
  - Plans to ensure participant confidentiality and anonymity
  - Plans for storage and handling of data
  - Information about what will happen to the data after the study
  - Information about how any data and images may be used
  - State whether it will be possible to identify any individuals

- ☐ Copies of any information sheets to be given to participants (include recruitment information (e.g. adverts, posters, letters)
- ☐ A copy of the participant consent form
- ☐ Copies of data collection sheets, questionnaires, etc.

*I confirm that all of the above is included in the application.*

*As the applicant, I confirm that I have read and understand the ethical requirements for my study and have read and complied with the University of Nottingham Code of Research Conduct and Research Ethics.*

**Signature of applicant** ..... **Date** .....

*As supervisor, I confirm that I have checked the details of this application.*

**Signature of supervisor** ..... **Date** .....

**NB: The signature of the supervisor on this part of the application DOES NOT indicate supervisor approval for expedited review. If supervisor approval is granted, then the front page of the application MUST be signed for approval to be confirmed.**

## REFERENCES

- Annett, J. (2002). Subjective rating scales: Science or art? *Ergonomics*, 45, 966–987.
- Ayoub, M.M. and Dempsey, P.G. (1990).
- Bainbridge, L. and Sanderson, P. (2005). Verbal protocol analysis. In: Wilson, J.R. and Corlett, N. (eds.), *Evaluation of Human Work*. London, U.K.: Taylor & Francis Group, pp. 159–184.
- Bisantz, A.M. and Drury, C.G. (2005). Applications of archival and observational data. In: Wilson, J.R. and Corlett, N. (eds.), *Evaluation of Human Work*. London, U.K.: Taylor & Francis Group, pp. 61–82.
- Blake, R. and Sekuler, R. (2006). *Perception*, 5th edn. New York: McGraw-Hill International.
- Cebola, N., Golightly, D., Wilson, J.R. and Lowe, E. (16–19 April 2012). On-call scheduling in rail maintenance – diary research challenges in an industrial context. *Proceedings of the IEHF Annual Conference*, Blackpool, U.K., pp. 328–335.
- Cranwell, J., Golightly, D., Fischer, J., Sharples, S. and O'Malley, C. (2012). Using mobile applications that combine self-report micro surveys to enhance GPS tracking data. *International Journal of Education and Psychological Assessment*, 11(1), 55–74.
- Dane, F.C. (1990). *Research Methods*. Pacific Grove, CA: Brooks Cole.
- Drury, C.G. (1995). Methods for direct observation of performance. In: Wilson, J.R. and Corlett, E.N. (eds.), *Evaluation of Human Work*, 2nd edn. London, U.K.: Taylor & Francis Group.
- Fechner, G.T. (1860). *Elemente der Psychophysik*. Leipzig, Germany: Breitkopf and Härtel.
- Ferguson, E. and Cox, T. (1993). Exploratory Factor Analysis: A Users' Guide. *International Journal of Selection and Assessment*, 1(2), 84–94.
- Field, A. and Hole, G. (2003). *How to Design and Report Experiments*. London, U.K.: Sage Publications Ltd.
- Fischer, S.L. and Dickerson, C.R. (2014). Applying psychophysics to prevent overexposure: On the relationships between acceptable manual force, joint loading, and perception. *International Journal of Industrial Ergonomics*, 44, 266–274.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233.
- Fostervold, K.I., Buckmann, E. and Lie, I. (2001). VDU-screen filters: Remedy or the ubiquitous Hawthorne effect? *International Journal of Industrial Ergonomics*, 27(2), 107–118.
- Garg, A., Waters, T., Kapellusch, J. and Karwowski, W. (March 2014). Psychophysical basis for maximum pushing and pulling forces: A review and recommendations. *International Journal of Industrial Ergonomics*, 44(2), 281–291.



- Garneau, C.J. and Parkinson, M.B. (2013). Considering just noticeable difference in assessments of physical accommodation for product design. *Ergonomics*, 56(11), 1777–1788.
- Gescheider, G.A. (1985). *Psychophysics: Method, Theory and Application*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gescheider, G.A. (1997). *Psychophysics: The Fundamentals*, 3rd edn. Mahweh, NJ: Lawrence Erlbaum Associates.
- Hignett, S. (2014). Commentary on the paper by Heimrich Kanis entitled: ‘Reliability and validity of findings in ergonomics research’. *Theoretical Issues in Ergonomics Science*, 15(1), 50–54.
- Igbaria, M., Schiffman, S.J. and Wieckowski, T.J. (1994). The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. *Behaviour & Information Technology*, 13(6), 349–361.
- Kanis, H. (2014). Reliability and validity of findings in ergonomics research. *Theoretical Issues in Ergonomics Science*, 15(1), 1–46.
- Kanis, H., Schoormans, J.P.L. and Green, W.S. (2014). Reliability and validity revisited. *Theoretical Issues in Ergonomics Science*, 15(1), 62–68.
- Karwowski, W., Sherehiy, B., Gaddie, P.T., Khalaf, T. and Quesada, P.M. (2007). The effects of lifting instructions on the psychophysically selected lifting load limits: A need for reappraisal. *Occupational Ergonomics*, 7(1), 43–51.
- Kline, P. (1994). *An Easy Guide to Factor Analysis*. London, U.K.: Routledge.
- Lee, J.-L., Stone, E.A., Wakabayashi, H. and Tochihara, Y. (2010). Issues in combining the categorical and visual analog scale for the assessment of perceived thermal sensation: Methodological and conceptual considerations. *Applied Ergonomics*, 41, 282–290.
- Li, K.W., Yu, R. and Han, X.L. (2007). Physiological and psychophysical responses in handling maximum acceptable weights under different footwear-friction conditions. *Applied Ergonomics*, 38(2007), 259–265.
- Liu, Y. (2003). Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. *Ergonomics*, 46(13–14), 1273–1292.
- Lowenthal, K.M. (1996). *An Introduction to Psychological Tests and Scales*. London, U.K.: UCL Press.
- Maikala, R.V., Dempsey, P.G. and Maynard, W.S. (March 2014). Honoring Dr. Vincent Matthew Ciriello special issue guest editorial. *International Journal of Industrial Ergonomics*, 44(2), 197–342.
- Nichols, S. (1998). Development of a questionnaire to measure attitudes towards virtual reality. In: Hanson, M. (ed.), *Contemporary Ergonomics*. London: Taylor & Francis Group, pp. 146–150.
- Nichols, S. (1999). Virtual Reality Induced Symptoms and Effects: Theoretical and methodological issues. PhD thesis, University of Nottingham, Nottingham, U.K.
- Nielsen, J. (2000). Why you only need to test with 5 users, Article posted on Nielsen Norman Group website: <http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>. Last accessed 5 May 2014.
- Nielsen, J. (2012). How many test users in a usability study? Article posted on Nielsen Norman Group website: <http://www.nngroup.com/articles/how-many-test-users/>. (Accessed 5 May 2014).
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Science Education*, 15, 625–632.
- Nunnally, J.C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Oppenheim, A.N. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*. London, U.K.: Pitman.
- Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L. and Morrisroe, G. (2005). The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload. *Applied Ergonomics*, 36(6), 681–693.
- Robson, C. (2002). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers*, 2nd edn. Oxford, U.K.: Blackwell Publishing.
- Rolland, J.P., Meyer, C., Arthur, K. and Rinalducci, E. (2002). Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence*, 11(6), 610–625.
- Ryan, B., Wilson, J.R., Sharples, S., Morrisroe, G. and Clarke, T. (March 2009). Developing a rail ergonomics questionnaire. *Applied Ergonomics*, 40(2), 216–229.
- Sinclair, M.A. 2005. Participative assessment. In: Wilson, J.R. and Corlett, N. (eds.), *Evaluation of Human Work*, 3rd edn. London, U.K.: Taylor & Francis Group, pp. 83–112.
- Sharples, S., Cobb, S., Moody, A. and Wilson, J.R. (2008). Virtual Reality Induced Symptoms and Effects (VRISE): Comparison of Head Mounted Display (HMD), desktop and projection display systems. *Displays*, 29(2), 58–69.

- Sharples, S., Millen, L., Golightly, D. and Balfe, N. (2011). The impact of automation in rail signalling operations. *Journal of Rail and Rapid Transit. Proceedings of IMechE, Part F*, 225(2), 179–191.
- Snook, S.H. (1978). The design of manual handling tasks. *Ergonomics*, 21, 963–985.
- Snook, S.H. and Ciriello, V.M. (1991). The design of manual handling tasks: Revised tables of maximum acceptable weights and forces. *Ergonomics*, 34(9), 1197–1213.
- Southall, D. (1985). The discrimination of clutch-pedal resistances. *Ergonomics*, 28(9), 1311–1317.
- Stanton, N.A. (2014). Commentary on the paper by Heimrich Kanis entitled 'Reliability and validity of findings in ergonomics research': Where is the methodology in ergonomics methods? *Theoretical Issues in Ergonomics Science*, 15(1), 55–61.
- Sun, X., Golightly, D., Cranwell, J., Bedwell, B. and Sharples, S. (2013). Participant experiences of mobile device-based diary studies. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 5(2), 62–83.
- Surdick, R.T., Davis, E.T., King, R.A., Corso, G.M., Shapiro, A., Hodges, L. and Elliott, K. (1994). Relevant cues for the visual perception of depth: Is where you see it where it is? *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, 38(1), 1305–1309.
- Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, 78(1), 71–80.
- Thurstone, L.L. and Chave, E.J. (1929). *The Measurement of Attitudes*. New York: Columbia University Press.
- Vink, P. (2014). Reaction on the paper 'Reliability and validity of findings in ergonomics research' of Kanis (2013). *Theoretical Issues in Ergonomics Science*, 15(1), 47–49.

---

# 5 Qualitative Methodology

*Sue Hignett and Hilary McDermott*

## CONTENTS

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| Introduction.....                                                  | 119 |
| What Is Qualitative Methodology?.....                              | 120 |
| A Brief Philosophy for Qualitative Methodology.....                | 120 |
| Doing a Qualitative Project.....                                   | 121 |
| Data Sources.....                                                  | 123 |
| Sampling .....                                                     | 123 |
| Data Management, Display and Analysis .....                        | 124 |
| Procedural Framework .....                                         | 129 |
| Validity and Reliability in Qualitative Methodology .....          | 131 |
| Respondent Validation.....                                         | 133 |
| Triangulation .....                                                | 133 |
| Critical Appraisal .....                                           | 133 |
| CAQDAS (Computer-Assisted Qualitative Data Analysis Software)..... | 134 |
| NVivo .....                                                        | 134 |
| Conclusion .....                                                   | 136 |
| References.....                                                    | 136 |

## INTRODUCTION

Qualitative methodology is increasingly used to lead and support ergonomics and human factors (E/HF) studies in a range of contexts. This chapter provides an insight into the practical use of qualitative methodology in E/HF and outlines the theory and principles which underpin the use of such methodologies. We present an overview of the main qualitative approaches and provide guidance on undertaking a qualitative project; to support this, we describe the use of thematic analysis, including examples of computer-aided analysis; discuss validity, reliability, and critical appraisal in relation to research design and analysis and introduce a qualitative data management software package (NVivo).

A review of the use of qualitative methodologies in E/HF (Hignett, 2001) found early examples of methodological exploration including exploratory sequential data analysis (ESDA, Sanderson and Fisher, 1997:1472). ESDA was described as ‘any empirical undertaking seeking to analyse systems, environmental and/or behavioural data (usually recorded) in which the sequential integrity of events has been preserved’ and was used as an umbrella term to group established techniques rather than proposing new ones. Sanderson and Fisher suggested three E/HF traditions supporting the exploration of qualitative methodologies (ESDA): (1) the behavioural tradition using directly observable laboratory-based experimentation; (2) the cognitive tradition to model indirectly observed or symbolic behaviour, for example in human-computer interaction and (3) the social or naturalistic tradition of social sciences as a more recent development.

The use of qualitative methodologies in E/HF has increased considerably since this chapter was first written in 2005, for example in inclusive design (Fisk et al., 2009), participatory ergonomics

(Dixon and Theberge, 2011), organisational ergonomics (Berlin, 2011) and workplace analysis (Lundh et al., 2011). This reflects the maturation of the qualitative–quantitative debate and the appreciation of alternative epistemological (ways of knowing) perspectives, giving new and critical insights for accepted practices (Symon and Cassell, 2004).

One worrying trend has been the use of qualitative methodologies without a clear statement (and understanding) of the underpinning philosophy. Although projects can be carried out to a satisfactory level without reference to the relevant theory (ontology and epistemology), this can lead to problems with the quality of the project. Silverman (2006) views this as a failure in analysis with, for example, an emphasis on the exploration of a problem and very limited testing of the findings (explanation) or proof that contrary evidence has been sought. This will be further discussed in this chapter to set out both a generic process for qualitative projects and mechanisms to ensure that reliability and validity are considered and addressed.

## WHAT IS QUALITATIVE METHODOLOGY?

Qualitative research is concerned with the understanding of meaning. Qualitative researchers are interested in how people make sense of their world and how they interpret and experience different events.

The key points in defining qualitative methodology are (Hignett, 2001; Robson, 2011)

- *Non-numerical*: A consistency in representing the world in terms of words and pictures rather than numbers. There is little or no use of numerical data or statistical analysis in qualitative research.
- *Scale*: Focuses on a few cases with many variables. Qualitative studies tend to follow an idiographic mode of inquiry by focussing on the individual rather than groups (nomothetic).
- *Sampling strategy*: Develops during the study (not pre-assigned) driven by an inductive logic as theoretical concepts and ideas emerge.
- *Iterative data collection and analysis*: The details of the procedure are not fixed in advance (flexible design), and the focus is liable to change during the study.
- *Context*: Situations are described from the perspective of participants to understand phenomena in context (natural settings).
- *Influence of the researcher*: Recognising the importance of the values of the researcher by reflecting on their interaction before and during the project. Objectivity is not valued as it may distance the researcher from the participants.

## A BRIEF PHILOSOPHY FOR QUALITATIVE METHODOLOGY

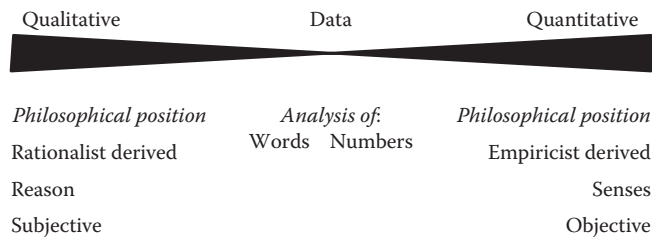
In order to fully understand and engage with the rest of this chapter, it is necessary to introduce and define some important key concepts related to scientific research.

*Ontology*: The nature of our world and what we believe exists for us to know.

*Epistemology*: The branch of philosophy which studies knowledge and is concerned with how and what can we know.

*Methodology*: A general approach to studying research topics; ‘method’ refers to a specific research technique (Silverman, 2006).

In the philosophy of science, there are two extreme poles (rationalism and empiricism) (Figure 5.1). A rationalist has the belief that reason is the primary source of knowledge, with certain innate ideas that exist in the mind prior to all experience. The rationalists are roughly grouped as taking a world view, or ontology, with human reason as the central tenet. In contrast, an empiricist believes that



**FIGURE 5.1** A simple qualitative–quantitative continuum. (From Hignett, S., Using qualitative methodology in ergonomics: Theoretical background and practical examples, PhD thesis, University of Nottingham, Nottingham, U.K., 2001.)

there is absolutely nothing in the mind that is not experienced through the senses. This dichotomy can be traced back to ancient Greek philosophy where the mind (reason, rationalism and subjectivism) was contrasted to matter (senses, empiricism and objectivism) (Murphy et al., 1998:15).

Symon and Cassell (2012) give an example using occupational stress to describe the different ontological positions. As an objectivist concept, stress is ‘a real phenomenon that exists and can be measured’. In contrast, the subjectivist description is that ‘stress does not exist in individuals. It has no real, independent status separate from the act of knowing and it is created through perceiving/ knowing the social world’.

There is an ontological realism in which the dichotomy of qualitative and quantitative methodologies can coexist with the ‘recognition of the existence of a real, independent world which operates according to natural necessity with a corresponding position of epistemological relativism’ (Bhaskar, 1975:250). Robson (2011) describes this as the *pragmatic approach* which recognises the existence of the physical world as well as the emergent social and psychological world. He describes knowledge as ‘being both constructed and based on the reality of the world experienced and lived in...[with]...current truth, meaning and knowledge as tentative and as changing over time’ (Robson, 2011:29). This allows different (even conflicting) perspectives to be included and described as part of the research analysis and interpretation.

*Methodology* refers to the choices made about the problem definition (topic) and methods (techniques) of data collection and analysis (Table 5.1). It is used as an umbrella term to indicate the theory and account of how research is, or should be, carried out. Mixing methodologies is very different to mixing methods as methodologies reflect their underlying philosophy, for example discourse analysis and thematic analysis. In order to use mixed methods, it is better to use a methodology which supports a middle ground ontology, for example realism (pragmatic approach). Table 5.1 offers a list (not exhaustive or mutually exclusive) and description of different qualitative methodologies that E/HF researchers/practitioners may consider applying.

The rest of this chapter refers to thematic-type analyses, which focus on the discovery of regularities and comprehension of meaning in data (e.g. what people think or feel), in contrast to looking at the characteristics of language as the ‘linguistic turn or rhetorical approach’ (Gomm et al., 2000:262). Some of the qualitative methodologies in Table 5.1 (e.g. discourse analysis and conversation analysis) use linguistic analysis as opposed to thematic analysis.

## DOING A QUALITATIVE PROJECT

This section looks at data sources, sampling, data collection, data management and data analysis. There are common (generic) processes across some stages of data collection and analysis which have been described as thematic in nature (Bannister et al., 1994; Braun and Clarke, 2006; Robson, 2011), but conclusion drawing reflects the underpinning methodological choices to interpret the findings within the context of the appropriate body of literature.

**TABLE 5.1**  
**Qualitative Methodologies**

| Methodology                                     | Description                                                                                                                                                                                                                                                                                                                                          |
|-------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Content analysis (quantitative and qualitative) | Textual analysis to compare, contrast and categorise data.<br>Quantitative: Establish a set of categories, then count the number of instances that fall into each category for statistical analysis.<br>Qualitative: Used to understand categories and see how these are used to describe, for example, working life.                                |
| Thematic analysis                               | Generic approach (foundation method) to the analysis of qualitative data. Can be used as a realist method to report experiences, meaning and the reality of participants, or as a constructionist method to examine the ways in which events, realities, meanings and experiences are the effects of a range of discourses operating within society. |
| Grounded theory                                 | Highly systematic research approach for the iterative collection and analysis of qualitative data using constant comparison and theoretical sampling until theoretical saturation is achieved.                                                                                                                                                       |
| Template analysis                               | Similar to grounded theory but starts with an initial coding template which is verified and/or modified through data collection/analysis. The template is used to show relationships between the themes (often as a hierarchy) and stops after reading texts three to four times.                                                                    |
| Ethnography                                     | An ethnograph is a description and interpretation of a cultural or social group or system. The product of an objective knowledge based on personal interaction and subjective experience.                                                                                                                                                            |
| Phenomenology                                   | Description of the experience of everyday life as it is internalised in the subjective consciousness of the individual. The structure and essence of the experience of a phenomenon for people.                                                                                                                                                      |
| Discourse analysis                              | Analysis of the process of communication as everyday language by exploring structure, function and patterns. Discourses 'prescribe' appropriate behaviours and attributes across a range of social domains.                                                                                                                                          |
| Conversation analysis                           | Detailed analysis of audio and audio-visual recordings of naturally occurring social interaction to identify the interactional practices used by speakers to produce their own conduct and to interpret and deal with the conduct of others.                                                                                                         |

*Sources:* Chenitz, W.C. and Swanson, J.M., *From Practice to Grounded Theory: Qualitative Theory in Nursing*, Addison-Welsey Publishing Company, Menlo Park, CA, 1986; Creswell, J.W., *Qualitative Inquiry and Research Design: Choosing among Five Traditions*, Sage Publications Ltd., Thousand Oaks, CA, 1998; Miles, M.B. and Huberman, A.M., *Qualitative Data Analysis: An Expanded Source Book*, 2nd edn., Sage Publications Ltd., Thousand Oaks, CA, 1994; Patton, M.Q., *Qualitative Evaluation and Research Methods*, 2nd edn., Sage Publications Ltd., Newbury Park, CA, 1990; Schwandt, T.A., *Qualitative Inquiry: A Dictionary of Terms*, Sage Publications Ltd., Thousand Oaks, CA, 1997; Strauss, A. and Corbin, J., *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Sage Publications Ltd., Newbury Park, CA, 1990; Silverman, D., *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*, 3rd edn., Sage Publications Ltd., London, U.K., 2006; Robson, C., *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*, 3rd edn., John Wiley & Sons Ltd., Chichester, U.K., 2011; Symon, G. and Cassell, C., Promoting new research practices in organisational research, in Cassell, C. and Symon, G., eds., *Essential Guide to Qualitative Methods in Organisational Research*, Sage Publications Ltd., London, U.K., 2004, pp. 1–10, Chapter 1; Symon, G. and Cassell, C., *Organizational Qualitative Research: Core Methods and Current Challenges*, Sage Publications Ltd., London, U.K., 2012; Braun, V. and Clarke, V., *Qual. Res. Psychol.*, 3, 77, 2006; King, N., Using templates in the thematic analysis of text, in Cassell, C. and Symon, G., eds., *Essential Guide to Qualitative Methods in Organisational Research*, Sage Publications Ltd., London, U.K., 2004, pp. 256–270, Chapter 21.

**TABLE 5.2**  
**Project Design Decisions**

|                             |                                                                                                                                                                                                                                                                     |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. What is studied          | Intellectual question, site, participants (data sources). Identify the range of solutions, devices or stratagems that can be used.                                                                                                                                  |
| 2. Under what circumstances | Access and entry to site and participants. Ethics.                                                                                                                                                                                                                  |
| 3. For what duration        | Time frame.                                                                                                                                                                                                                                                         |
| 4. Research strategy        | Methodology including personal position and viewpoint (conceptual framework) with respect to the research question, site and participants. Theory of scientific knowledge (philosophy) with assumptions about the nature of reality and the role of the researcher. |
| 5. Methods                  | Preference for certain methods, e.g. watching, questioning, listening, reading.                                                                                                                                                                                     |
| 6. Procedural framework     | A systematic sequence of procedural steps to give a good audit trail showing how data were collected and then managed with respect to the analysis.                                                                                                                 |

*Sources:* Janesick, V.J., The dance of qualitative research design: Metaphor, methodolatry and meaning, Chapter 2, in Denzin, N.K. and Lincoln, Y.S. (eds.), *Strategies of Qualitative Inquiry*, Sage Publications Ltd., Thousand Oaks, CA, 1998, pp. 35–55; Silverman, D., *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*, 3rd edn., Sage Publications Ltd., London, U.K., p. 15, 2006.

When planning a qualitative project, there are basic research design decisions to be made (Table 5.2). The research strategy (point 4) is a key qualitative defining dimension. One way of achieving this is to use a conceptual framework to identify implicit and explicit theories and relationships. In a qualitative project, this will be an ongoing activity, so a project diary can be useful to record memos, mind maps, reflective thoughts and iterations throughout the project. This diary can also be used as part of the audit trail in establishing reliability.

## DATA SOURCES

Although qualitative data sources are many and varied, there are three basic types: spoken (interviews), visual (observations) and written data (documents) as shown in Table 5.3.

## SAMPLING

The sampling strategy for any research project should be defensible with respect to the appropriate relationship (or logic) of the sample and the intellectual question. Sampling is one of the key dimensions in defining qualitative methodology. Non-probability samples are commonly used in qualitative work, whereas quantitative methodology uses probability sampling to test (or falsify) a pre-existing theory (Saunders, 2012:39).

The range of sampling strategies is grouped into similar logics (Table 5.4) for

- Spreading the net
- Following up leads
- Focussing
- Analysis

It is usually necessary to use more than one sampling logic during a qualitative project (Sandeslowski et al., 1992). The sampling strategy will develop during the project, so this might start by spreading the net, and then go on to following up leads or focussing on a specific characteristic and conclude with some form of analysis sampling.

The size of sample for qualitative studies is ambiguous and dependent on the question being investigated, theoretical saturation and credibility (Saunders, 2012:44). Some authors have suggested minimum sizes for non-probability samples as 5–25 (interviews, Kvale and Brinkman, 2009),



**TABLE 5.3**  
**Qualitative Data Types**

| <b>Miles and Huberman (1994)</b>                              | <b>Wolcott (1992)</b> | <b>Dingwall (1997)</b>                                     | <b>Mason (2002)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------------------------------------------------------------|-----------------------|------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Interviews<br>(spoken data)                                   | Asking                | Asking questions<br><br>Researcher–researched relationship | Ontology – people’s knowledge, views, understandings, interpretations, experiences and interactions are meaningful properties of the social reality which the research questions are designed to explore.<br><br>Epistemology – a legitimate way to generate data on these ontological properties is to interact with people, to talk to them, to listen to them and to gain access to their accounts and articulations.<br><br>Knowledge and evidence are contextual, situational and interactional; so each interview will be different, reflexive and responsive to the situation, context and interaction.                                                                                                                     |
| Observation<br>(visual data)                                  | Watching              | ‘Hanging out’<br><br>Transactions between members          | Ontology – a data collection method which sees interactions, action and behaviours, and the way people interpret these and act on them, as central.<br><br>Epistemology – the knowledge or evidence or the social world can be generated by observing, participating in, or experiencing natural or real-life settings, interaction situations and so on, based on the premise that these kinds of settings, situations and interactions reveal data, and that it is possible for the researcher to be an interpreter, or knower of such data as well as an experienced observer or participant observer.                                                                                                                          |
| Documentary<br>(written data),<br>visual and<br>other methods | Examining             | Reading the papers                                         | Ontology – (1) written words, texts, documents, records, objects, visual or spatial phenomena or aspects of social organisation, shape, form, etc. are meaningful constituents of the social world in themselves; (2) interest in the processes by which they are produced or consumed; (3) belief that they act as some form of expression or representation of relevant elements of the social world or (4) that aspects of the social world can be traced or read through them.<br><br>Epistemology – texts, documents, written records, visual documents, visual records, objects, artefacts and phenomena or visualisation (as a process more than a thing) can provide or count as evidence of these ontological properties. |

20–35 (Grounded Theory, Cresswell, 1998), 20–30 (Template Analysis, King, 2004) and 4–12 for a homogeneous population or 12–30 for a heterogeneous population (Kuzel, 1992).

## **DATA MANAGEMENT, DISPLAY AND ANALYSIS**

Table 5.5 sets out a generic process for the management of qualitative data for a thematic analysis (Marshall and Rossman, 1989; Dey, 1993; Miles and Huberman, 1994:10; Sanderson and Fisher, 1997; Braun and Clarke, 2006). All start in the same way by organising, reducing and describing the data through primary coding.

**TABLE 5.4**  
**Sampling Strategies**

| <b>Sampling</b>                               | <b>Purpose</b>                                                                                                                                                                                                                                                   |
|-----------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Spreading the net</i>                      |                                                                                                                                                                                                                                                                  |
| Purposive                                     | Maximum variation/open sampling. Picking a wide range of variation on dimensions of interest (time, location, events and people) to provide the greatest opportunity to gather the most relevant heterogeneous data about the phenomenon.                        |
| Mixed purposeful                              | Triangulation, flexibility, meets multiple interests and needs.                                                                                                                                                                                                  |
| Convenience (haphazard)                       | Save time, money, effort. Poorest rationale, lowest credibility. Yields information-poor cases.                                                                                                                                                                  |
| <i>Following up leads</i>                     |                                                                                                                                                                                                                                                                  |
| Theoretical                                   | Analyst jointly collects, codes and analyses the data and then decides which data to collect next and where to find them in order to develop and inform the theory as it emerges.<br>Central tenet of grounded theory (secondary or analysis sampling strategy). |
| Snowball (volunteer)                          | Identifies cases of interest from people who know people, who know what cases are information-rich, i.e. good examples for study.                                                                                                                                |
| Opportunistic                                 | Following new leads during field work, taking advantage of the unexpected flexibility.                                                                                                                                                                           |
| <i>Focussing</i>                              |                                                                                                                                                                                                                                                                  |
| Homogenous                                    | Focuses, reduces variation, simplifies analysis using one sub-group (e.g. occupation or level in organisation)                                                                                                                                                   |
| Typical case                                  | Illustrates what is typical, normal, average, trying to find more than one case.                                                                                                                                                                                 |
| Intensity                                     | Information-rich cases that manifest the phenomenon intensively but not extremely, such as above/below average.                                                                                                                                                  |
| Stratified purposeful                         | Illustrates characteristics of a particular sub-group of interest, facilitates comparisons.                                                                                                                                                                      |
| <i>Analysis sampling (inductive analysis)</i> |                                                                                                                                                                                                                                                                  |
| Extreme or deviant case                       | Learning from highly unusual manifestations of the phenomenon of interest. Qualify findings and specify variations or contingencies in the main patterns observed.                                                                                               |
| Confirming and disconfirming cases            | Elaborating and deepening initial analysis, seeking exceptions, looking for variations. Disconfirming cases limit conclusions and indicate points of greatest variation.                                                                                         |
| Criterion                                     | Picking cases that need some criterion, such as children abused in a treatment facility. Quality assurance.                                                                                                                                                      |
| Multiple case                                 | Grounding a finding using replication strategy.                                                                                                                                                                                                                  |
| Indiscriminate                                | Choosing sites, persons and documents that will maximise opportunities for verifying the storyline, relationships between categories and for filling in poorly developed categories.                                                                             |

*Source:* Developed from Patton, M.Q., *Evaluation and Research Methods*, 2nd edn., Sage Publications Ltd., Newbury Park, CA, p. 182, 1990; Miles, M.B. and Huberman, A.M., *Qualitative Data Analysis: An Expanded Source Book*, 2nd edn., Sage Publications Ltd., Thousand Oaks, CA, 1994; Kuzel, A.J., Sampling in qualitative inquiry, in Crabtree, B.F. and Miller, W.L., eds., *Doing Qualitative Research*, Sage Publications Ltd., Newbury Park, CA, 1992; Coyne, I.T., *J. Adv. Nurs.*, 26, 623, 1997; Strauss, A. and Corbin, J., *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Sage Publications Ltd., Newbury Park, CA, 1990; Glaser, B., *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*, Sociology Press, Mill Valley, CA, 1978; Yin, K., *Applications of Case Study Research*, Cosmos Corporation, Washington, DC, 1991; Saunders, M.K., Choosing research participants, in Symon, G. and Cassell, C., eds., *Organizational Qualitative Research: Core Methods and Current Challenges*, Sage Publications Ltd., London, U.K., 2012, pp. 35–52, Chapter 3.

**TABLE 5.5**  
**Levels of Data ‘Reading’ (Analysis)**

|              |                                                                                                                                                                                                                                                                                                                                                                   |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Literal      | Form, content, structure, style, layout, etc.<br>Interview – words and language used, sequence of interaction, form and structure of dialogue and literature content.                                                                                                                                                                                             |
| Interpretive | Construction or documentation of a version of what the researcher thinks the data means or represents or what can be inferred from it. Reading through or beyond the data for texts, artefacts, visual images, etc. Includes interviewee’s and researcher’s interpretations and understandings or versions and accounts of how sense is made of the social world. |
| Reflexive    | Locates the researcher as part of the data that has been generated and explores the researcher’s role and perspective in the process of generating and interpreting the data.                                                                                                                                                                                     |

Source: Mason, J., *Qualitative Researching*, 2nd edn., Sage Publications Ltd., London, U.K., 2002, p. 149.

Spoken and visual data may be converted into written data via transcription but can also be analysed directly as audio and image files. Silverman (2006:204) lists the benefits of recording data as providing a public record, allowing repeat reviews/analyses (and improving transcriptions) and preserving the sequences of talk. Transcriptions require careful planning and attention to detail (Barbour, 2008). Although no ‘standard’ has been established, it is important to decide what level of transcription is required. For linguistic analysis, a greater level of exactness will be required (including, for example, pauses and changes in pitch), whereas a thematic analysis may only require a verbatim transcription. Transcribing is a time-consuming process but does allow familiarisation (immersion) with the data at an early stage (Grey, 2009). If an external transcription service is used, the researcher will need to listen to the original empirical dataset (audio or video files) and check the transcription for accuracy.

Step two marks the start of the analysis by displaying the data either visually or as text. Step three mostly involves interpretation or conclusion drawing. As the qualitative process is iterative, these steps are intertwined and cyclical, rather than linear. Theoretical saturation is the point during a research project at which no new instances are being identified and data collection can cease. It developed from grounded theory but is now more widely used in contemporary qualitative projects for directing the sampling and data collection phases.

There are different levels of written data from field notes to verbatim transcription; the level and type of detail will be driven by the research question, methodology, resources and planned analysis. The data can be read literally, interpretively and/or reflexively (Table 5.5). The levels are not exclusive, and many research projects will include analysis on all three levels. Typically, thematic analysis does not require the same level of detail as, for example, conversation and discourse analysis (Robson, 2011).

Primary analysis starts by generating initial coding categories within the data before moving to identifying themes and trends. A code is a label which exemplifies the same theoretical or descriptive idea and is applied to ‘chunks’ of data of varying sizes (single words, sentences, paragraphs, sections of pictures and segments of audio). An example of initial coding can be found in Table 5.6.

Robson (2011:479) provides an extensive list of coding options which can include specific acts, behaviours, events, activities, strategies, practices, tactics, states, meanings, participation (involvement), relationships, conditions or constraints, consequences, settings and reflexive (researcher’s role).

Coding at this initial stage can be considered as a data reduction exercise and a data organisation process in preparation for more in-depth and detailed analyses. However, it can also be used as an analytical strategy as a more subtle process of having ideas and using concepts to describe the

**TABLE 5.6**  
**An Example of Initial Coding of Qualitative Data**

|                                                                                                                                                                                                                                                                                       |                                                                                              |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| <i>P: Usually people in this job love the job, and they love going in the mountains so they'll go anyway, and also the work ethic, I think you'll find people in this job have a pretty strong work ethic, not the kind of people who would phone in sick.</i>                        | <i>People love the job</i><br><i>People love the mountains</i><br><i>Strong work ethic</i>   |
| <i>Int: Do you think there is reflective practice? Acknowledging that this isn't doing you any good?</i>                                                                                                                                                                              |                                                                                              |
| <i>P: Yes, people aren't stupid, possibly if you work for yourself, there's not much choice. I don't know what the case is for getting insurance working self-employed in the mountaineering industry; I would imagine the cost is quite high, so a lot of people haven't got it.</i> | <i>People are aware of what they are doing</i><br><i>High cost of insurance – preventive</i> |
| <i>Int: Some people have also mentioned the benefits of outdoor activity?</i>                                                                                                                                                                                                         |                                                                                              |
| <i>P: I suppose it would ensure you have a good strong heart, but I think the wear and tear on the joints is the main factor.</i>                                                                                                                                                     | <i>Benefits of outdoor work</i><br><i>Drawbacks of outdoor work</i>                          |

data by noticing relevant phenomena collecting examples of those phenomena, and analysing the phenomena in order to find commonalities, differences, patterns and structures. Tools for reducing data include

- Contact summary sheets are used to summarise thoughts and to review the data immediately after data collection as a quality assurance or reflective mechanism. They are also used in analysis for identifying key points and areas which need further exploration.
- Memoing during transcribing, reading and re-reading of the data. This is a continuation of reflection which started when formulating the conceptual framework. Memos are ideas about codes and their relationships as they strike the researcher during coding (Glaser, 1978).

Pattern coding is the second level, where coding is used to expand, transform and reconceptualise the data (Coffey and Atkinson, 1996). Miles and Huberman (1994) suggested ways of using pattern coding:

- To map the codes by network displays (tables, maps and models) to show how components interact.
- To check out the codes in the next wave of data collection to ensure that all the coded extracts and entire data set are included. This is likely to require a revision of the initial codes.

Step two in Table 5.7 is data display. Miles and Huberman (1994), Robson (2011) and Braun and Clarke (2006) provide a number of suggestions for data display including

1. Context chart, where the inter-relationships between roles and groups are mapped in graphic form.
2. Checklist matrix to tabulate the data in terms of a specific question.
3. Time-ordered display to show the flow and sequence of events. This is similar to an activity record or critical incident chart.
4. Conceptually ordered display to show well-defined themes and interactions.
5. Thematic map as tables and mind maps to explore the relationships both between codes and themes, and the different levels of themes.

**TABLE 5.7**  
**Generic Data Management Steps**

| Author(s)                   | Step One: Data Reduction                                                                                                                                                                                                                                                          | Step Two: Data Display                                                                                                                                                                                                                                                                                    | Step Three: Conclusion Drawing/Verification                                                                                                                                                                             |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Miles and Huberman (1994)   | Summarising and packaging the data, managing field notes, transcripts, etc.<br>Data are reduced in anticipatory ways as conceptual frameworks are chosen, and cases and questions are refined.<br>Data are summarised, coded and broken down into themes, clusters or categories. | Repackaging and aggregating the data, e.g. matrices, charts, graphs, networks.<br>Data display describes diagrammatically pictorial or visual forms in order to show what those data imply to give an organised, compressed assembly of information that permits conclusion drawing and/or action taking. | Developing and testing propositions to construct an explanatory framework, regularities, patterns, explanations, causal flows.<br>Conclusion drawing and verification using different tactics, e.g. analytic induction. |
| Marshall and Rossman (1989) | Organising the data.                                                                                                                                                                                                                                                              | Generalising categories, themes and patterns.                                                                                                                                                                                                                                                             | Testing emergent hypotheses against the data, looking for contradictions.<br>Searching for explanations for the data.<br>Writing the report.                                                                            |
| Dey (1993)                  | Describing, including context of action, intentions and process of social action.                                                                                                                                                                                                 | Classifying, as themes and codes, to give meanings.                                                                                                                                                                                                                                                       | Connecting concepts.                                                                                                                                                                                                    |
| Braun and Clarke (2006)     | Familiarisation: transcribing, reading and re-reading data to note initial ideas.<br>Generating initial codes and collating data relevant to each code.<br>Collating codes into themes.                                                                                           | Reviewing coded data within themes and across whole dataset by generating a thematic map of the analysis.<br>Defining and naming themes to refine the specifics and overall story.<br>Generating clear definitions and names for each theme.                                                              | Producing the report.<br>Selection of vivid, compelling extract examples, final analysis of examples relating back to the research questions and literature.                                                            |
| Sanderson and Fisher (1997) | Commenting, chunking and coding.                                                                                                                                                                                                                                                  | Connecting and converting.                                                                                                                                                                                                                                                                                | Comparing, constraining and computing.                                                                                                                                                                                  |

Data display can also be used as part of the analysis, to identify and find relationships and then test these against the data. This creates the framework for the next process of analytic induction as part of the conclusion drawing.

Conclusion drawing is listed as step three, whereas, in fact, there is a fuzzy boundary between this and the previous step. Analytic induction is the process whereby negative or extreme cases are sought to firstly test, secondly extend the scope and finally determine the limits of the developing theory. Basically the theory is revised until all the exceptions are eliminated by inclusion (Fielding and Fielding, 1986:89; Mason, 2002:136; Silverman, 2006:295). At this stage, it is important to bring all the reflective strands together. Testing the interpretation will include checking against researcher biases as acknowledged at the start of the project in the conceptual framework, as well as the influences (underpinning philosophy) of the methodology.

## PROCEDURAL FRAMEWORK

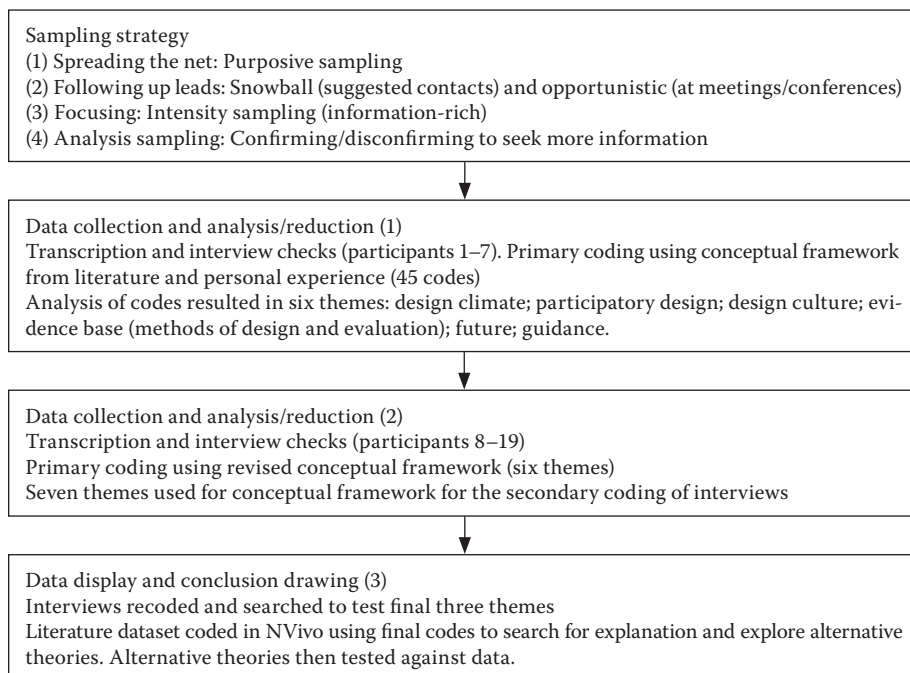
This section will give examples of how thematic analysis was used in two E/HF studies.

### Example 5.1

The first example is a study that explored the use of building design guidance by health-care architects and planners in the United Kingdom (Hignett and Lu, 2009). Sixteen participants were interviewed individually or as a group with the interviews being audio-taped and transcribed verbatim for analysis. Contact summary sheets were completed after each interview to capture immediate thoughts and summarise the main points (Miles and Huberman, 1994). The sampling strategy followed a four stage process involving (1) purposive sampling with a wide range of experts, (2) snowball and opportunistic sampling to follow new leads during field work, (3) intensity sampling to obtain clarification on aspects of the interpretation and (4) confirming and disconfirming case sampling to elaborate and deepen the analysis, seek exceptions, limit conclusions and look for variations.

The verbatim transcripts were returned to participants for confidentiality, and accuracy checking and data from one interview (a group interview of three participants) was requested to be withdrawn from the study (Figure 5.2).

Initially, the data were organised and reduced using the qualitative data management tool in NVivo (see later in this chapter) (Bazeley and Richard, 2000) and classified into 45 preliminary codes. Following this, a detailed secondary coding was conducted within the codes to identify six higher level themes. The first five interviews (participants 1, 2 + 3, 4, 5 + 6, 7) were re-read and re-coded with these higher level themes. During the coding of the next two interviews (participants 11, 12 + 13), minor changes were made to clarify and expand the definition of each theme. As the study progressed, theoretical saturation was achieved. The analysis of the final four interviews



**FIGURE 5.2** Procedure framework. (From Hignett, S. and Lu, J., *Appl. Ergon.*, 40, 608, 2009.)

(participants 14 + 15, 16, 17 + 18, 19) developed the codes to more inclusive descriptions and explanations resulting in three final themes:

1. Design culture including design climate and participatory design
2. Research/evidence including design history, international research and quality issues
3. Future guidance needs including concepts and philosophy and patient expectations

The use of a qualitative approach for this project enabled a reflexive position to be maintained with respect to both the research question and the data. The analytical induction strengthened the development of the interpretation by searching for negative cases in the data and the analysis and by reviewing the themes until an acceptable representation was produced.

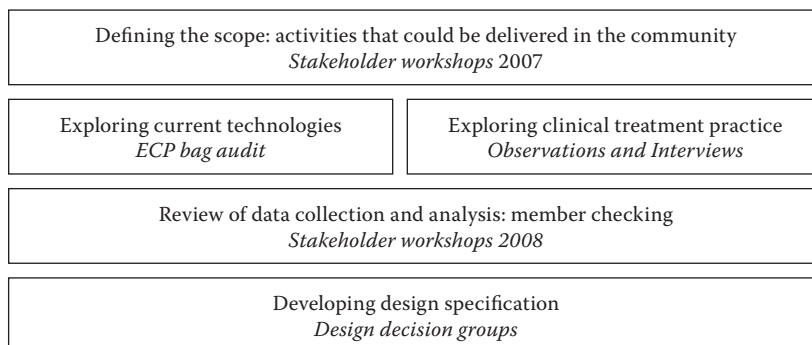
### Example 5.2

This study describes an explorative and descriptive study about urgent care technologies from the acute, primary care and ambulance sectors with 125 staff and 88 patients over 18 months (Hignett et al., 2011). Qualitative data were collected from stakeholder workshops, portable technology audit, treatment observations in emergency departments and walk-in centres, and design decision groups (DDGs) (see Chapters 11 and 34 on inclusive design and participatory ergonomics) (Figure 5.3).

The scope of the project was defined at two stakeholder workshops in 2007 with participants recruited from different health-care sectors (acute, ambulance and primary care) using mixed purposive sampling to give variation on the dimensions of interest. The use of current technologies was explored with an audit of portable equipment and consumables for emergency care practitioners (ECPs) selected using homogenous sampling. Data were collected using interviews and a checklist and then grouped for comparison with the equipment lists. Clinical treatment practices for urgent complaints were explored through 84 observations of patient treatment (stratified purposeful sampling). Observational (link analysis) and staff interview data (hierarchical task analysis) were collected until theoretical saturation was achieved and no additional information was being generated.

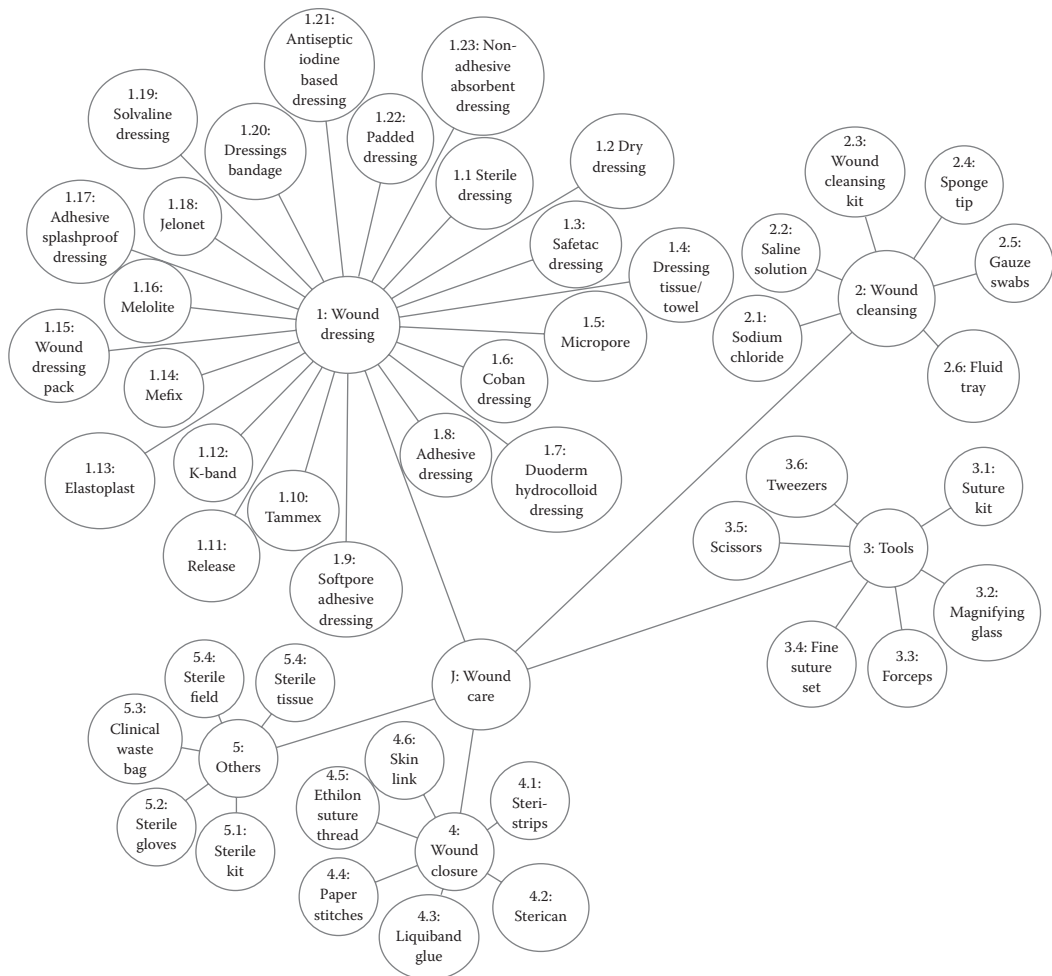
In 2008, a second workshop was held to present the findings of the 2007 workshops, audit and observations. Data were collected as a series of semi-structured questions in individual workbooks. The triangulated data from the four datasets (stakeholder workshops in 2007 and 2008, ECP bag audit and observations) were reviewed by the second researcher before being taken to the DDGs. This resulted in eight primary themes: wound care (Figure 5.3); drugs, gases and vaccines; diagnostic equipment; office stationery; hygiene/sanitation; additional technologies (e.g. razor); immobilisation (e.g. splint/sling) and phlebotomy equipment/consumables (Figure 5.4).

The final stage used criterion sampling for the DDGs as a quality assurance approach to test the ongoing analysis. Data were collected using round robins, word maps and drawing exercises



**FIGURE 5.3** Urgent care technologies. (From Hignett, S. et al., *Emerg. Med. J.*, 28, 192, 2011.)





**FIGURE 5.4** Example of triangulation of four datasets for wound care. (From Hignett, S. et al., *Emerg. Med. J.*, 28, 192, 2011.)

to produce both paper mock-ups and prototypes for the discussion and modification of the design requirements. From these iterations, it emerged that the supporting portable technologies for urgent care should consist of a three level technology system.

## VALIDITY AND RELIABILITY IN QUALITATIVE METHODOLOGY

The words ‘validity’ and ‘reliability’ are taken from quantitative methodology (see Chapter 1) and need interpretation in a qualitative context. Table 5.8 summarises some of the terms used by Guba and Lincoln (1981), Lincoln and Guba (1985), Miles and Huberman (1994) and Robson (2011).

Internal validity addresses issues of credibility and authenticity in the research (Table 5.9). At an operational level, this can be established through the use of an audit trail and the analytic induction process of testing theory. External validity looks at issues of generalisability and transferability. Reliability addresses the issues of auditability or quality control. This could be the consistency by which instances are assigned to the same code in analysis, or on a broader level to the wider process itself.

**TABLE 5.8**  
**Alternative Terms for Validity and Reliability**

| Internal Validity | External Validity | Reliability   |
|-------------------|-------------------|---------------|
| Credibility       | Fittingness       | Auditability  |
| Truth value       | Applicability     | Consistency   |
| Trustworthiness   | Transferability   | Dependability |
| Authenticity      | Generalisability  |               |

**TABLE 5.9**  
**Strategies to Establish Validity and Reliability**

| Internal Validity (Credibility)                         | External Validity (Generalisability) | Reliability (Auditability)              |
|---------------------------------------------------------|--------------------------------------|-----------------------------------------|
| Richness of descriptions                                | Description of original sample       | Clarity of research question            |
| Comprehensiveness of account                            | Sampling limitations                 | Role and status of researcher           |
| Linkage to theory                                       | Scope, boundaries and variation      | Meaningful analysis across data sources |
| Internal coherence of concepts                          | Transferability potential            | Clear connection to theory              |
| Evidence of addressing areas of uncertainty (surprises) | Data presentation                    | Scope of data collection                |
| Negative evidence                                       | Member checking                      | Coding checks                           |
| Checking the meaning of outliers and extreme cases      | Further testing                      | Data quality checks                     |
| Rival explanations                                      | Replication                          | Peer review                             |
| Member checking                                         |                                      |                                         |
| Accuracy of conclusions and predictions                 |                                      |                                         |

*Source:* Robson, C., *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*, 3rd edn., John Wiley & Sons, Chichester, U.K., 2011.

The detail given with respect to the context, researcher bias, sampling strategy and history of the research question can all help to establish the conditions whereby the findings could be transferred to another setting. Dingwall (1997:62) gave three tests for general validity:

1. Distinguish clearly between data and analysis.
2. Examine the extent to which the study has looked for contradictory or negative evidence and set out to test statements proposed on theoretical grounds or reported from previous studies.
3. See how it reflects the interactive character of social life and deals even-handedly with the people being studied.

At a fundamental level, the aim of all research should be to convince the reader of the validity, reliability and relevance of the research findings. Whether this is achieved using large sample sizes and statistical tests, or by detailed descriptions of a situation or point of view, depends on the design of the investigation or exploration. If the reader is able to use the research by incorporating the findings in their own work, then boundaries have been extended and knowledge has been generated, and robust scientific research has been achieved.

## RESPONDENT VALIDATION

Respondent validation, also known as member checking, is when the interpretation of the researcher is presented back to the participants as part of the conclusion drawing and verification (Walker, 1989). This is a different process to accuracy checking of data where an interview transcript is returned to the interviewee for that purpose. Mays and Pope (1995) suggested that member checking could be used to add to both the internal (authenticity check) and the external validity (transferability of findings).

## TRIANGULATION

Triangulation is another method which can be used to establish both internal and external validity. It refers to the use of more than one data source, method or investigator and the convergence of these to add credibility to a study (see Chapter 1). The underlying rationale behind this is that through combining different methods, the weakness of one method may be addressed through the strength of another (Magnusson et al., 2005).

Coffey and Atkinson (1996:14) argue that if the philosophical (ontological) positions have not been defined, the combination of different analyses would look as if they had been stuck ‘together like children’s building blocks in order to create a single edifice’, resulting in a comparison which will be between optimal and inferior methods and data. If the data are generated from different contexts, Silverman (2006:295) suggests that triangulation may be inappropriate as it may ignore the ‘context-based and skilful characters of social interactions’.

## CRITICAL APPRAISAL

Table 5.10 shows a critical appraisal question set from Hignett et al. (2003) incorporating criteria for assessing quality from Robson (2011:486). This can also be used to internally review a project prior to writing the final report to identify the strengths and limitations of individual project designs.

---

**TABLE 5.10**

**Question Set to Critically Appraise Qualitative Studies**

1. Is the research question/aim/objective of the study clearly described?
  2. Are the research methods appropriate to the question being asked?
  3. Was the qualitative method that was used made clear in the aims of the study?
  4. Is there a clear connection to an existing body of knowledge/wider theoretical framework?
  5. Is the context for the research adequately described and accounted for?
  6. Are the criteria for, and approach to, sample selection, data collection and analysis described clearly and systematically applied?
  7. Does the paper describe the sample in terms of gender, ethnicity, social class etc. (if appropriate)?
  8. Was the sample appropriate?
  9. Were the processes of fieldwork and the means of data collection described adequately?
  10. Is the relationship between the researcher and the researched considered and have the latter been fully informed?
  11. Is sufficient consideration given to how findings were derived from the data and how the validity of the findings was tested (negative examples, member checking)?
  12. Has evidence for and against the researchers interpretation been considered?
  13. Are the findings systematically reported and is sufficient original data reported to justify conclusions?
-

## CAQDAS (COMPUTER-ASSISTED QUALITATIVE DATA ANALYSIS SOFTWARE)

The use of CAQDAS helps to make nonlinear research processes more systematic by adding both flexibility and rigour. Most CAQDAS packages can handle a wide range of data types including text, video, audio and graphics and have tools to enable content searching, linking, coding, querying, memoing, modelling, mapping and networking (Sinkovics and Alfoldi, 2012). It is important to reinforce the point which is often made when discussing the use of computers in data analysis; although they can help with the organisation and processing of the data, computers cannot do the thinking, interpreting or relationship exploring: this must come from the human (researcher).

### NVivo

NVivo is a qualitative data management software package that is designed to store and manage different types of qualitative data (Richards, 2009; Daws, 2011), for example: interview transcripts; field notes (from observations); focus group transcripts; literature review notes; audio and video sources; other types of documents (e.g. minutes, diagrams); and any type of *external* data (e.g. documents that cannot be imported).

An NVivo project has two fundamental parts:

1. Sources (data) are created or imported into the project as data are collected.
2. Nodes are created for themes or categories by coding sources.

It can be used to systematically organise and manage data, code and retrieve data (Figure 5.5), search and theorise data using queries, record new ideas and annotations as *memos*, modify and merge *nodes*, draw *models* to represent ideas about the project (can be linked directly to the data/node), develop an *audit trail* of searches and analyses to enhance the research reliability, and use *visualisations* (e.g. word trees and tag clouds) to gain new/different perspectives on the data.

In NVivo, the themes/categories and coding are stored at nodes. Nodes provide a location for gathering all related material in one place to look for emerging patterns and ideas. Nodes can



FIGURE 5.5 Importing sources into NVivo (version 9).

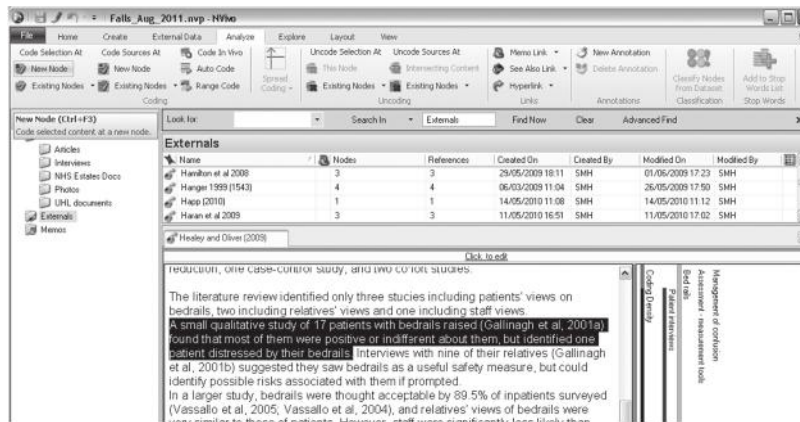


FIGURE 5.6 Coding in NVivo.

be created and named as new nodes (in vivo), from the words in the data; created from prior ideas by coding 'down' (theoretical framework/research question) or created and coded 'up' from meanings in the data (Figure 5.6).

Nodes can be cut and pasted between trees (themes) and can also be merged to organise nodes if it is found that the contents contain the same ideas or concepts. As the analysis progresses, the *find* and *query* functions can be used to explore patterns and validate theories. Outputs include tag clouds (Figure 5.7), models (static, Figure 5.3, and dynamic) and coding/project reports.

These outputs support the analysis by providing a systematic record of the data relevant to each stage as a log (or audit) trail of the project, lists of project items (content and coding), formal reports on the analysis (coding structure) and patterns of analysis in models and visualisations to see different views of the data.



FIGURE 5.7 Visual representation of coding as tag cloud (no indication of context or interpretation of data).

## CONCLUSION

This chapter sets the scene for using qualitative methodologies in ergonomics research. There has been an ongoing philosophical debate between two poles (represented here as qualitative–quantitative) for over 2000 years. We have discussed how, in E/HF, it is possible to take a middle ground position and combine both qualitative and quantitative methodological approaches for the same research question. In taking this pragmatic position, a generic process for doing qualitative research has been described using a thematic analysis approach with three steps of data reduction, data display and conclusion drawing.

At the moment, it seems (from conference proceedings and journal publications) that ergonomics is currently more at the quantitative end, but there are influences tipping the balance back towards the qualitative side.

## REFERENCES

- Bannister, P., Burman, E., Parker, I., Taylor, M. and Tindall, C. (1994) *Qualitative Methods in Psychology: A Research Guide*. Buckingham, U.K.: Open University Press.
- Barbour, R. (2008) *Introducing Qualitative Research*. London, U.K.: Sage Publications Ltd.
- Bazeley, P. and Richards, L. (2000) *The NVivo Qualitative Project Book*. London, U.K.: Sage Publications Ltd.
- Berlin, C. (2011) Human factors experiences in context – comparing four industrial cases using a soft systems framework. *The Ergonomics Open Journal*, 2011(4): 131–144.
- Bhaskar, R. (1975). A realist theory of science. Leeds: Leeds Books. *Reclaiming Reality: A Critical Introduction to Contemporary Philosophy*. London, U.K.: Verso.
- Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3: 77–101.
- Chenitz, W.C. and Swanson, J.M. (1986) *From Practice to Grounded Theory: Qualitative Theory in Nursing*. Menlo Park, CA: Addison-Welsey Publishing Company.
- Coffey, A. and Atkinson, P. (1996) *Making Sense of Qualitative Data*. Thousand Oaks, CA: Sage Publications Ltd.
- Coyne, I.T. (1997) Sampling in qualitative research. Purposeful and theoretical sampling: Merging or clear boundaries? *Journal of Advanced Nursing*, 26: 623–630.
- Creswell, J.W. (1998) *Qualitative Inquiry and Research Design: Choosing among Five Traditions*. Thousand Oaks, CA: Sage Publications Ltd.
- Daws, L. (2011) Learning NVivo9. Kihi Consultancies. info@kihi.com.au. <http://www.kihi.com.au/> (accessed 19 Nov 2014).
- Dey, I. (1993) *Qualitative Data Analysis: A User Friendly Guide for Social Scientists*. London, U.K.: Routledge.
- Dingwall, R. (1997) Accounts, interviews and observations. In Miller, G. and Dingwall, R. (eds.), *Context and Method in Qualitative Research*. London, U.K.: Sage Publications Ltd.
- Dixon, S.M. and Theberge, N. (2011) Contextual factors affecting task distribution in two participatory ergonomic interventions: A qualitative study. *Ergonomics*, 54(11): 1005–1016.
- Fielding, N.G. and Fielding, J.L. (1986) *Linking Data*. Qualitative research method series 4. London, U.K.: Sage Publications Ltd.
- Fisk, A.D., Rogers, W.A. and Charness, N. (2009) *Designing for Older Adults: Principles and Creative Human Factors Approaches*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Glaser, B. (1978) *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*. Mill Valley, CA: Sociology Press.
- Gomm, R., Needham, G. and Bullman, A. (2000) *Evaluating Research in Health and Social Care*. London, U.K.: The Open University/Sage Publications Ltd.
- Grey, D. (2009) *Doing Research in the Real World*, 2nd edn. London, U.K.: Sage Publications Ltd.
- Guba, E.G. and Lincoln, Y.S. (1981) *Effective Evaluation*. San Francisco, CA: Jossey-Bass Publishers.
- Hignett, S. (2001) Using qualitative methodology in ergonomics: Theoretical background and practical examples. PhD thesis, University of Nottingham, Nottingham, U.K.
- Hignett, S., Crumpton, E., Ruzsala, S., Alexander, P., Fray, M. and Fletcher, B. (2003) *Evidence-Based Patient Handling: Tasks, Equipment and Interventions*. London, U.K.: Routledge.



- Hignett, S., Jones, A. and Bengner, J. (2011) Portable treatment technologies for urgent care. *Emergency Medicine Journal*, 28: 192–196.
- Hignett, S. and Lu, J. (2009) An investigation of the use of health building notes by UK Healthcare Building Designers. *Applied Ergonomics*, 40: 608–616.
- Janesick, V.J. (1998) The dance of qualitative research design: Metaphor, methodolatry and meaning, Chapter 2. In Denzin, N.K. and Lincoln, Y.S. (eds.), *Strategies of Qualitative Inquiry*. Thousand Oaks, CA: Sage Publications Ltd., pp. 35–55.
- King, N. (2004) Using templates in the thematic analysis of text, Chapter 21. In Cassell, C. and Symon, G. (eds.), *Essential Guide to Qualitative Methods in Organisational Research*. London, U.K.: Sage Publications Ltd., pp. 256–270.
- Kuzel, A.J. (1992) Sampling in qualitative inquiry. In Crabtree, B.F. and Miller, W.L. (eds.), *Doing Qualitative Research*. Newbury Park, CA: Sage Publications Ltd.
- Kvale, S. and Brinkman, S. (2009) *Interviews: Learning from the Craft of Qualitative Research Interviewing*. Thousand Oaks, CA: Sage Publications Ltd.
- Lincoln, Y.S. and Guba, E.G. (1985) *Naturalistic Inquiry*. Beverly Hills, CA: Sage Publications Ltd.
- Lundh, M., Lützhöft, M., Rydstedt, L. and Dahlman, J. (2011) Working conditions in the engine department – a qualitative study among engine room personnel on board Swedish merchant ships. *Applied Ergonomics*, 42(2): 384–390.
- Magnusson, C., Finnerty, G. and Pope, R. (2005) Methodological triangulation in midwifery educational research. *Nurse Researcher*, 12(4): 30–39.
- Marshall, C. and Rossman, G.B. (1989) *Designing Qualitative Research*. Thousand Oaks, CA: Sage Publications Ltd.
- Mason, J. (2002) *Qualitative Researching*, 2nd edn. London, U.K.: Sage Publications Ltd.
- Mays, N. and Pope, C. (July 1, 1995) Reaching the parts other methods cannot reach: An introduction to qualitative methods in health and health sciences research. *BMJ*, 11: 42–45.
- Miles, M.B. and Huberman, A.M. (1994) *Qualitative Data Analysis: An Expanded Source Book*, 2nd edn. Thousand Oaks, CA: Sage Publications Ltd.
- Murphy, E., Dingwall, R., Greatbatch, D., Parker, S. and Watson, P. (1998) Qualitative research methods in health technology assessment: A review of the literature. *Health Technology Assessment*, 2(16): iii–ix, 1–274.
- Patton, M.Q. (1990) *Qualitative Evaluation and Research Methods*, 2nd edn. Newbury Park, CA: Sage Publications Ltd.
- Richards, L. (2009) *Handling Qualitative Data: A Practical Guide*, 2nd edn. London, U.K.: Sage Publications Ltd.
- Robson, C. (2011) *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*, 3rd edn. Chichester, U.K.: John Wiley & Sons Ltd.
- Sanderson, P.M. and Fisher, C. (1997) Exploratory sequential data analysis: Qualitative and quantitative handling of continuous observational data, Chapter 44. In Salvendy, G. (ed.), *Handbook of Human Factors and Ergonomics*, 2nd edn. New York: John Wiley & Sons, pp. 1471–1513.
- Sandeslowski, M., Holditch-Davis, D. and Harris, B.G. (1992) Using qualitative and quantitative methods: The transition or parent-hood of infertile couples. In Gilgun, J.F., Daly, K. and Handel, G. (eds.), *Qualitative Methods in Family Research*. Newbury Park, CA: Sage Publications Ltd., pp. 301–323.
- Saunders, M.K. (2012) Choosing research participants, Chapter 3. In Symon, G. and Cassell, C. (eds.), *Organizational Qualitative Research: Core Methods and Current Challenges*. London, U.K.: Sage Publications Ltd., pp. 35–52.
- Schwandt, T.A. (1997) *Qualitative Inquiry: A Dictionary of Terms*. Thousand Oaks, CA: Sage Publications Ltd.
- Silverman, D. (2006) *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*, 3rd edn. London, U.K.: Sage Publications Ltd.
- Sinkovics, R.R. and Alfoldi, E.A. (2012) Facilitating the interaction between theory and data in qualitative research using CADQAS, Chapter 7. In Symon, G. and Cassell, C. (eds.), *Organizational Qualitative Research: Core Methods and Current Challenges*. London, U.K.: Sage Publications Ltd., pp. 109–131.
- Strauss, A. and Corbin, J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage Publications Ltd.
- Symon, G. and Cassell, C. (2004) Promoting new research practices in organisational research, Chapter 1. In Cassell, C. and Symon, G. (eds.), *Essential Guide to Qualitative Methods in Organisational Research*. London, U.K.: Sage Publications Ltd., pp. 1–10.
- Symon, G. and Cassell, C. (2012) *Organizational Qualitative Research: Core Methods and Current Challenges*. London, U.K.: Sage Publications Ltd.



- Walker, M. (1989) Analysing qualitative data: Ethnograph and the evaluation of medical education. *Medical Education*, 23: 498–503.
- Wolcott, H.F. (1992) Posturing in qualitative inquiry. In LeCompte, M.D., Milroy, W.L. and Preissie, J. (eds.), *The Handbook of Qualitative Research in Education*. New York: Academic Press, pp. 3–52.
- Yin, K. (1991) *Applications of Case Study Research*. Washington, DC: Cosmos Corporation.

---

# 6 Task Analysis

*Andrew Shepherd*

## CONTENTS

|                                                      |     |
|------------------------------------------------------|-----|
| Introduction.....                                    | 140 |
| Tasks.....                                           | 141 |
| Tasks in Experiments .....                           | 141 |
| Tasks in the Real World.....                         | 142 |
| The Issue of Definition.....                         | 143 |
| Variations in Task Analysis Method.....              | 144 |
| Tasks in the Working Environment .....               | 144 |
| Developments in Task Analysis Methodology .....      | 146 |
| Work-Study .....                                     | 146 |
| Sensorimotor Charts.....                             | 146 |
| Miller's Task Description and Task Analysis .....    | 147 |
| Increase of 'Cognitive' Tasks.....                   | 148 |
| Hierarchical Task Analysis .....                     | 149 |
| Cognitive Task Analysis .....                        | 151 |
| Task Classification and Tabular Task Analysis.....   | 152 |
| HTA as a Task Analysis Framework .....               | 153 |
| Types of Plan.....                                   | 155 |
| Fixed Sequences .....                                | 155 |
| Cued Actions .....                                   | 155 |
| Contingent Fixed Sequence.....                       | 156 |
| Choices .....                                        | 156 |
| Cycles.....                                          | 156 |
| Time-Sharing.....                                    | 158 |
| Combining Plan Elements and Plans .....              | 158 |
| Redescription Rather than Decomposition.....         | 158 |
| Stopping Rules .....                                 | 158 |
| An Illustration .....                                | 159 |
| Operations, Goals and the Re-Scoping of Levels ..... | 159 |
| Inheritance of Context.....                          | 159 |
| Bringing the Task into Focus.....                    | 160 |
| Examining Behaviour Appropriate to Operations.....   | 160 |
| Validating a Task Analysis .....                     | 160 |
| Conclusions.....                                     | 161 |
| Keith Duncan: An Appreciation.....                   | 161 |
| References.....                                      | 162 |

## INTRODUCTION

Task analysis refers to a set of methods used to identify characteristics of work situations that influence human performance. Effective task analysis enables research findings and insights concerning work behaviour to be applied in a focussed, principled and justifiable way to improve or evaluate human performance, thereby promoting greater system reliability and effectiveness.

Applied human factors work is generally undertaken at the instigation of a client or by addressing a problem of practical concern to an interested party. Such clients possess views, standards, values and constraints about the task, all of which need to be taken into account, because they may influence how a task is or should be carried out. Judgement about psychological or physiological task factors should also take account of the context in which the task is undertaken.

Task analysis should ensure that judgements and decisions made are relevant to a client's requirements. If a client (or agent) agrees that the task analysis reflects an acceptable account of the task, then this benefits the relationship between client and investigator, and supports effective communication between members of a design or management team. It contributes to a clearer agreement about where to focus or what to avoid. Thus, a task analysis promotes both validity of ergonomics and human factors (E/HF) work and clarifies the working relationship between client and investigator.

Addressing real problems and issues in work situations means identifying what people are required to do at work. For the present account, it is convenient to refer to the worker as the 'operator' – someone who carries out actions, operates or maintains equipment, communicates or collaborates with others, processes information, provides support for others, supervises or manages processes and teams. Thus, 'operator' refers here to anybody who carries out operations on behalf of an organisation; it does not imply any particular type of work or any status.

The challenge to the investigator examining work for a particular purpose is to avoid distorting what the client believes is appropriate. A task analysis method should help the investigator to identify problems and propose ways of improvement.

There are many methods of task analysis and associated work-related methods of analysis. These are often not reconciled with each other. Indeed, one of the most noticeable features about the field of task analysis is that many investigators use the same terminology but refer to different things. It would be impossible to do justice to the vast range of task analysis approaches within this single chapter, and the reader is encouraged to refer to two handbooks: Kirwan and Ainsworth's *A Guide to Task Analysis* (1992a), which addresses task analysis through the many methods that have been applied to practical problems in industry, and Diaper and Stanton's *Handbook of Task Analysis for Human Computer Interaction* (2004). There are other reviews or collections of task analysis worth referring to (e.g. Stammers and Shepherd, 2005; Stanton, 2006) as well as chapters presented in human factors or human resource management textbooks.

The many disagreements about task analysis and explanations for why there are so many variants hinge in part on the fact that different investigators work in different domains. Different industrial and commercial domains contain different concerns and different patterns of work activity. Many proponents of *cognitive task analysis* (CTA) (see Diaper and Stanton, 2004) were prompted by consideration of users of different computer systems and applications, where environments are often more benign than other industrial contexts and recovery from error is more easily built into the functionality of the applications. This contrasts to people working on issues of human error and risk, in high-hazard process plants, the domain of focus of Kirwan and Ainsworth. Moreover, different investigators favour or accept different ways of construing behaviour and this can affect how data is collected and organised. Despite the differences, it is fair to say that every work analysis method has proven to have some practical benefit to somebody, somewhere.

This chapter aims to reconsider some fundamental issues of task analysis. It considers the concept of 'task' from the perspective of human factors even though, surprisingly, the concept is rarely discussed. It also considers how a task fits into the working environment. The account emphasises the breadth of factors that influence performance of tasks and which warrant attention

by investigators. Some of the key developments in task analysis methodology will then be outlined. Finally, it deals with *task analysis* as a framework for making decisions about *tasks* that can affect the way jobs are performed.

## TASKS

One reason for the varied approaches to task analysis is that investigators rarely define the word 'task'. One of the few definitions is Miller's (1962) – '... task is a group of discriminations, decisions and effector activities related to each other by temporal proximity, immediate purpose and a common man-machine output'. This excludes things that occur at diverse times, so, by this definition, it would be inappropriate to treat *management* or *supervision* as a task, because they oversee general goals\* and extend over long periods. It would rule out system supervisory tasks such as process control, air traffic control or nursing where the operator links different activities across long time-periods – when a system fails, its diagnosis and recovery planning often use information gleaned over previous hours, days or weeks.

Sometimes authors use 'task' to represent human behaviour – this is how Miller uses the word. This prompts concern, because it implies that a specific description of behaviour is appropriate to any given task, that everybody undertaking a task deploys the same psychological processes as other equally competent people. But people possess different behavioural traits, different knowledge about the world, different value systems and different experiences. Each person may adapt to the task requirements in different ways. Their strategies may be different, but their competence and error patterns may still be equivalent and acceptable. Cognition is generally private and it would be counterproductive to require everybody to think in the same way, even if it were possible or practical to know how each person thinks. Moreover, operators may apply different strategies on different occasions to deal with what may appear to be similar circumstances. For management, it may not matter how a goal is achieved, provided staff performance is appropriate, consistent and reliable.

Other authors such as Duncan (1972, 1974) treat 'task' in terms of the *purpose* of the work. Considering tasks in terms of purpose allows for variations in behaviour and avoids much unwarranted speculation concerning contentious attempts to explain behaviour.† Treating tasks in this way enables better agreement between client and investigator.

If tasks are prescribed sensibly and then carried out properly, this should contribute to a better outcome for the system of which the task is a part. Thus, task *analysis* aims to discover problems encountered or propose improvements to work methods. This operational conception of task suggests the task is a challenge that an operator is required to confront.

## TASKS IN EXPERIMENTS

Experimental psychologists prescribe tasks in precise and focussed ways such that performance variations can be attributed to controlled influences (see Chapters 1 and 2 for further discussion of experiments). For example a learning experiment may be undertaken to establish how a particular factor influences ease of learning. This factor, for example the amount of time the learner has to study material, may be varied so that the amount of learning achieved following a shorter study period can be compared with learning achieved following a longer study period. In order to attribute differences in amount learned to the length of the study period and not to some other extraneous

---

\* The word 'goal' is used extensively in discussions of task analysis. *Goal* can be used in a *hypothetical* psychological sense to imply a personal motivation for action, or it can be used in a *system* sense to refer to the target state of the system as instructed, for example by management. In this respect, an operator might be said to have a *goal*, because it is what they have been told to do as part of their 'contract' of employment. Whether this actually becomes their *motivation* is a matter for speculation. *Goal* is used in this chapter in this *system* sense.

† 'Understanding' behaviour remains an important aim for psychological research, but failure to understand behaviour fully is no impediment to effective applied work.

variable, the experimental situation is *controlled* so that other factors that could influence retention are equivalent for both conditions. The experimental method acknowledges there are many other things that might influence retention including group influences, layout of written material, age, gender and distraction. On another occasion, the experimenter might choose to examine the influence of one of these other variables, say, gender, and add ‘time to study the material’ to the list of items that require control in the basic experimental situation.

The experimental *task* is defined in terms of what participants are required to do and the conditions in which they are required to perform. Changes in the participants’ given target or in-task conditions means a different challenge – that is, a change in the task.

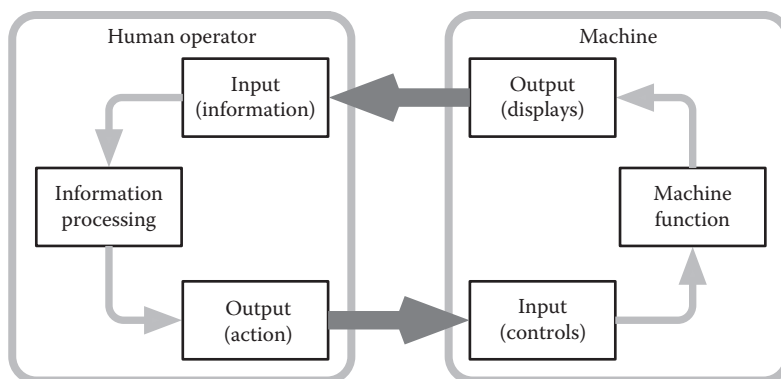
## TASKS IN THE REAL WORLD

Applied human factors investigators must deal with whatever is presented rather than try to constrain the task situations as in an experiment. Just as an experimenter tries to identify those aspects of the experimental situation in order to ‘control’ their influence in the experiment, an investigator of a real work situation must attempt to identify factors that could influence performance. Failure to do this may result in falsely rejecting a viable work solution, because its benefits were masked by something that had been overlooked.

Irrespective of its complexity, any task presents the operator with the following:

|             |                                                                                                                                                    |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Information | To prompt activity, enable appraisal of system status and make decisions                                                                           |
| Resources   | With which to effect change, including limbs, tools, energy, equipment, controls and language to interact with colleagues and articulate responses |
| Standards   | Constraining limitations on choice of action, set by management, the law or by convention                                                          |
| Constraints | Limiting how responses can be executed, imposed by management, the law, conventions and the affordances of the physical world                      |
| Environment | Affecting the capacity to think and to act, and the effects that actions will have on the controlled environment                                   |

A task viewed from this perspective can be represented as an interaction between a human operator and a ‘machine’\* over which the operator is required to exert some control or influence. Figure 6.1 shows a typical representation of this interaction. The operator obtains, then processes, information



**FIGURE 6.1** Typical representation of human–machine interaction.

\* The ‘machine’ can be an item of equipment, a plant, a colleague, a team, a vehicle, a customer, a computer, etc. or any combination of these things. It is a clumsy word to use in this respect, but so are the alternatives.

concerning the state of the machine in order to select actions aimed at exercising appropriate control. 'Machine function' determines how the machine then responds according to its own dynamics (e.g. via physical, chemical, legal and biological functions, or 'custom and practice').

'Machine function' is only a small box, but it is 'the door in the back of the wardrobe' opening onto the world. If we consider the context of air-traffic control (i.e. the air-traffic control system is our 'machine' in this context), an icon – the set of pixels representing an aircraft on the air-traffic controller's radar screen – is influenced by what is happening in the sky. The position of the aircraft depicted on the screen is influenced by how the computer is programmed to configure information sent to it via signals from the aircraft about the aircraft's position. The *actual* position of the aircraft depends upon conformity to the flight plan affected by the aircraft's performance, which is affected by payload, mechanical condition and punctuality of take-off. The position of the aircraft is further influenced by the performance of other aircraft encountered en-route, since these influenced previous air-traffic control instructions that caused the aircraft's flight path to be altered within current and previous air-space through which the aircraft has flown, affecting both time for the manoeuvres and fuel usage. It is possible that the apparent position of the aircraft is misleading due to malfunctioning of the communications or computing systems. These factors all influence the present set of circumstances and may need to be taken into account when an air-traffic control officer plans how to instruct pilots to change heading and altitude in order to avoid risk of collision with other aircraft. They may, therefore, all be things to be identified through a task analysis.

The aircraft icon on the screen could equally be controlled by simulation of the airspace with the icon's position controlled by a computer model of air-traffic movements. If this simulation conforms wholly to reality, then a trainee controller will experience the same as the real situation – as far as the controller is concerned, the task is the same. In practice, attaining total fidelity is extremely unlikely, especially with regard to the experience of stress and anxiety and the perception of risk associated with erroneous performance (see Chapter 8).

Real tasks may be focussed and local. Sometimes they entail information flowing directly from the environment to the senses, such as in craft skills where decisions taken affect how worked material will change as a consequence of action; the resultant change is monitored through feedback, enabling further adjustment in order to achieve the desired outcome. This can also apply to many computer application skills and interpersonal skills.

However, many tasks in human factors projects depend on instrumentation or computer interfaces for the operator to obtain information to make decisions. Even though interfaces are designed and built for the convenience of the user, information flow is influenced by the wider world. Thus, information may be presented intermittently, not at all, in a burst, infrequently or unreliably. These are all variations that affect expectation and judgement, as well as learning and experience.

Control actions may not directly influence the 'machine', but work through stages to cause delay, affecting machine function and feedback to the operator. This may cause significant *lag* between operator action and effect on the 'machine' and, thereby, affect performance. There may be no clear evidence to judge whether an action has been appropriate; factors in the real world may interact with each other and mislead the operator; in organisational settings, messages may be misinterpreted and acted upon inappropriately. 'Machine' states change according to principles governing their operation, according to environmental factors such as temperature, humidity, wear-and-tear and in accordance with other actions on the 'machine' by other people. Thus, the investigator needs to consider many peripheral things to obtain a rich picture of the impact of the task on the operator. All of these factors may be the concern of task analysis.

## THE ISSUE OF DEFINITION

Definitions of 'task' and 'task analysis' are routinely overlooked in expositions of task analysis. It may be assumed that there is a general consensus or that the ideas are made explicit by subsequent description of a task analysis method.

Following the guidance of Duncan (1972), *task* and *task analysis* are seen, for the purpose of this chapter, in terms of their *purpose* rather than speculation about psychological processes. *Task* is treated as a requirement to act, using available tools, equipment, plant, people and processes in order to achieve a required outcome, according to satisfactory standards and in accordance with prevailing situational constraints. *Task analysis* is considered the process of obtaining information about a working context that influences the performance effectiveness or leads to performance improvements. The task analyst may, therefore, pursue any justified lines of enquiry, including exploring the functionality and reliability of equipment, organisational and social factors and cognition, and use any appropriate method for obtaining information, such as observation, interviewing, questionnaires, checklists, protocol analysis, inspection of records and simulation.

## VARIATIONS IN TASK ANALYSIS METHOD

Some differences in task analysis methods are due to the different intentions of different analysts. Some set out ways to minimise work activity, or design an interface, or structure a work team. Different information may be required in these different circumstances, so different ways of seeking information may be used. Human performance is affected by

- Operator skills
- Motivation and fitness
- Machine information from which to make decisions
- Controls and resources to respond to varying circumstances
- The extent of collaboration with colleagues
- The extent to which work demands are prompted by productivity requirements, emergency situations or affected by inappropriate management
- Costs and consequences of errors

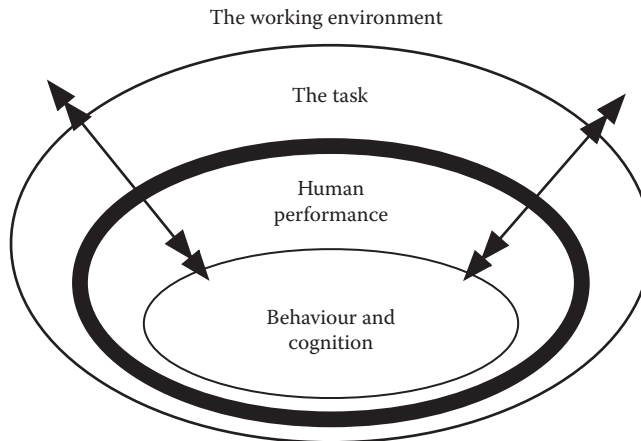
This is a substantial range of influences, some of which are more significant than others in different work domains and with respect to different problems.

Devising a task analysis method is rarely the main motivation for doing E/HF work. Usually, investigators are motivated to solve work problems or design improvements and require a systematic way of obtaining data to support their effort. Evolution of task analysis methods is often a by-product of such effort. Presented with this variety, someone needing a task analysis method has to select an approach they think will be appropriate. This may then be adapted, and another method is born. A collection of task analysis methods, such as that provided by Kirwan and Ainsworth (1992b), provides people who need such a method to select one according to the features of the required job. This can be a satisfactory solution; however, it presupposes that the analyst's intuitions about which tools to select are correct. Given that task analysis is used to find out about tasks, then prejudging the analytical tools to be used is likely to bias the analysis to construe tasks in a stereotyped way and fail to appreciate unexpected features. Task analysis should not oversimplify the world to favour a particular analytical method.

## TASKS IN THE WORKING ENVIRONMENT

Human performance resides within and interacts with a working environment. Carrying out a task requires controlling a working environment in order to achieve a desired outcome in terms of profit, safety, return on investment, environmental protection, benefits to personnel and a capacity to cope better with future events. It could mean all of these things. It is not the responsibility of an E/HF investigator to dictate what this purpose should be and what standards should be observed, although the investigator is at liberty to discuss these issues and contribute ideas, or decline work with a client whose requirements are unacceptable.





**FIGURE 6.2** A schematic of the working and task environment in conjunction with human operation.

There are several influences between the behaviour and cognition governing human performance and the working environment, as in Figure 6.2 (a simplified and task-focused interpretation of the model shown in Figure 6.2 in Chapter 1).

Information, communication and control flow in both directions. The working environment provides events that the human operator must deal with in achieving given goals, and the operator makes decisions to select and regulate action in order to influence the working environment. These events and the practice of dealing with them affect behaviour and cognition through priming and expectation; they influence choice and help prioritise action. Events also enable experiential learning to influence performance in the future. As events are dealt with, changes occur to the working environment and to the capabilities of a person undertaking the work.

Human response to events is controlled by behaviour and cognition, including capabilities arising from practice, memory, skill, pattern recognition, reasoning, aptitude, habit, and inference – even inspiration, instinct or intelligent guessing based on similarities or logical relationships between new and familiar events. For the purpose of effective system performance, it may not matter which psychological mechanisms are deployed, provided performance is effective and reliable. Equally, any improvement to the interaction between human performance and the working environment is beneficial in achieving system goals. Benefits from an E/HF intervention accrue from:

- Properly *specifying* how human operators should contribute to system performance
- ‘Filtering’ and representing information to the operator, hence *human–machine interface* design
- *Training* to improve behaviour and cognition and promote effective information processing
- Appropriate *personnel selection*
- *Mitigating* the risks associated with operator error to create a safer workplace and a safer context for learning and development
- Better team *organisation* to enable team members to collaborate for their common purpose and avoid occupational health problems due to workload
- Better *organisation, design and reliability of industrial plant and systems* that influence the frequency of demand for operator action

These are all reasons for using task analysis to identify information to inform E/HF decisions. Thus, information must often be obtained from the wider system and judgements made on how these factors affect tasks. It is not necessary to understand everything about a work situation, because many

features are not pertinent to how people carrying out tasks will perform. A task analysis method should enable the analyst to *focus* on those aspects of the task that are relevant to carrying out the task.

## DEVELOPMENTS IN TASK ANALYSIS METHODOLOGY

Work-study methods in the early part of the twentieth century were important, because industry depended upon rapid and reliable manual skill. As the century progressed, different forms of work organisation, made possible by developing technology, placed greater emphasis on cognitive skill. Equally, psychological aspects of analysing work assumed greater prominence when it was increasingly recognised that limitations of human performance were often due to central cognitive processes rather than physical or physiological factors. More recently developments in information and communications technology have impacted on work in terms of increased automation and computer tasks.

A full history of task analysis could go back a long way. Ancient civilisations were sophisticated in their thinking about weaponry, ensuring that weapons be manipulated effectively by their users as well as paying attention to their cutting edge. Vegetable plots in monastery gardens enabled kneeling monks to reach and work on the far side of a garden bed without damaging nearer plants. Signals at the end of railway tunnels, canal tunnels or mineshafts, or beacons to warn of foreign invasion, were all designed to alert people. Expert judgement to indicate when bread was cooked or swords suitably quenched, all entailed actions to obtain, and then rules to interpret, information. The twentieth century onwards may be the era of human factors and task analysis, but it is evident that the design and construction of mills, ships, weapons, furniture and buildings reflects purposeful analysis and design across the centuries.

## WORK-STUDY

Gilbreth (1912) devised a system of units to codify action, enabling work activity to be recorded in a convenient shorthand way. Work descriptions could then be used to consider efficient ways to carry out tasks; thus, work-study emerged as a discipline for observing and organising efficient ways to accomplish tasks (e.g. Currie, 1977). However, these approaches related to narrow criteria of success judged in terms of focusing on meeting and exceeding production targets at the expense of wider criteria, such as adaptability, attention, cognition, occupational health and job satisfaction. Work-study engineers focused effort on recording observable actions, because it was these that directly influenced production. These methods tended to be applied to short-cycle repetitive tasks, such as assembly tasks and bricklaying, effective organisation of which provided the greatest benefits for reducing work-cycle times and waste.

## SENSORIMOTOR CHARTS

Crossman (1956) introduced a broader psychological perspective, formally acknowledging human information processing within the cycle of work. This approach recorded cycles of skill relating observable actions and psychological processes such as *perception*, *planning* and *memory*. Of particular importance in this approach is the appreciation of cognitive processes in the mediation of action. The recognition of decision-making and judgement was part of an increasing trend as manufacturing technology moved from supporting short-cycle repetitive tasks through batch and continuous processing involving a higher degree of automation, to wholly integrated factory systems. This saw greater reliance on human operators monitoring partially self-regulating systems, then making decisions and planning strategies to rectify deviation from target.

A practical variation of this approach, widely used in industrial training at the time, was *skills analysis* developed by Seymour (1966) and this supported ways of carrying out repetitive tasks to help people to become skilled more quickly.

## MILLER'S TASK DESCRIPTION AND TASK ANALYSIS

A common approach to incorporating psychological factors in task analysis entails modelling human information processing. From the early 1950s, R.B. Miller, one of the most influential figures in the history of task analysis, published a number of articles setting out principles and guidance for analysing tasks. This work shows how the military, both in the United States and other countries, stimulated interest in task analysis as part of human factors in system development. Conflict and the threat of conflict have substantially influenced human factors development, because such events created requirement to prepare large numbers of people to operate increasingly sophisticated systems where the consequences of failure were dire. Moreover, it was necessary to achieve manning of military systems in a short time and to a high standard, providing impetus to manpower planning, task specification, equipment and interface design, training design, personnel selection and system evaluation.

Miller's approach, reported in numerous articles and technical reports, consisted of two main stages: *task description* and *task analysis* (e.g. Miller, 1966, 1967). In Miller's terms, a *task description* is a statement of what an operator must do expressed in *systems* or *operational* terms to change the state of the system.

In developing task descriptions, Miller drew attention to how tasks can be described hierarchically, such that operations or actions can be described in terms of sets of more detailed operations and so on. This is helpful, because tasks can then be treated as large or small activities. Some writers have attempted to classify different levels of focus; for example Diaper (1989) proposed a four-level hierarchy where 'activities' sit within 'subtasks' which sit within 'tasks' which sit within 'projects'. This categorisation might suit particular investigators working within specific work domains on particular problems, but it is less helpful as a general scheme. It begs the question about distinguishing between levels and discounts the possibility that additional levels might be equally useful.

In Miller's scheme, *task description* is followed by a phase called *task analysis* where the analyst explores tasks as *behavioural* entities. For this, Miller proposes an information processing model, presented graphically in Figure 6.3, entailing stages through which signals pass from the task environment through stages of *reception of task information*, *interpretation and problem solving* to *motor response mechanisms* dealing with the manner in which the response is selected (through *interpretation and problem solving*) in order to effect a change in the environment. Supporting this performance is *retention of task information* including the current status of specific parameters, as well as learned rules and principles governing decision-making. Influencing all activities is the

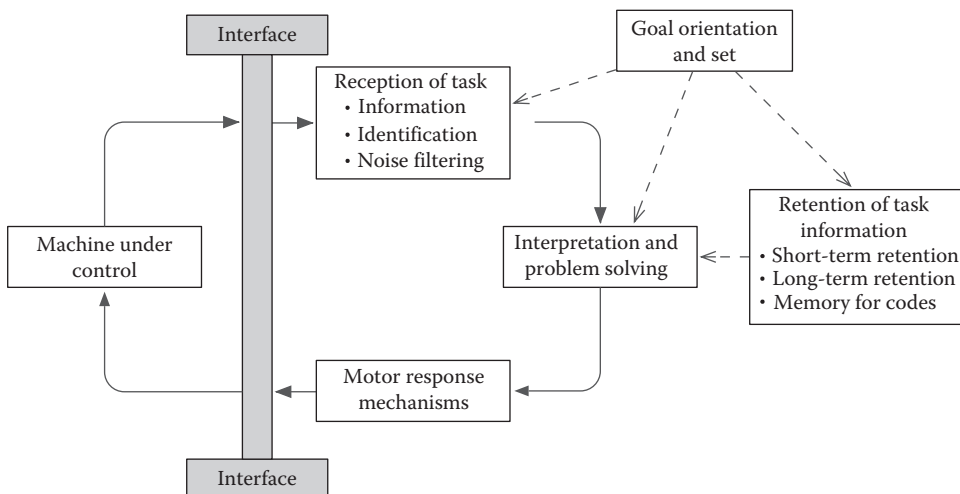


FIGURE 6.3 Schematic representation of Miller's scheme for task analysis.

*goal orientation and set*. The representation in Figure 6.3 is similar to many such human–system interaction models presented in E/HF (e.g. Wickens et al., 2013).

Miller's schematic representation is a theoretical model and open to debate. Such models are essentially *hypothetical*. Cognitive psychology offers justification for the components of such models, but it struggles to justify an acceptable universal view complying with all psychological challenges now, in the past and the future. This was of no concern to Miller who maintained that task analysis should be pragmatic. The hypothetical nature of task analysis was secondary to helping make useful suggestions – a common position in many applications of task analysis where investigators have used models of behaviour that are not necessarily wholly valid, but assist the process of systematic examination of tasks.

This approach is consistent with what was later to be referred to as *tabular task analysis*. Indeed, many investigators in areas such as human error analysis especially (e.g. Embrey, 1986; Kirwan, 2005) adopt a similar approach using taxonomies, often derived from a cognitive modelling approach, as a systematic basis for identifying potential sources of human error. Such approaches are useful, because they use a set of principled categories to examine systematically the different parts of task performance. Thus, the result of the task analysis is likely to be consistent, with regard to interpretation by a single analyst and also between two or more analysts working together on a large assignment. This is important for auditing systems to provide assurances of safety, especially in large hazardous engineered systems such as railways or power plants.\*

Distinguishing between *task description* and *task analysis*, as Miller proposed, may be useful in tasks where people carry out a number of self-contained actions. Here, it is possible to describe distinct activities according to Miller's *categorical* task description phase that has been agreed with a client, then explore them in turn using a task analysis approach to make hypotheses about cognition. In many tasks, however, component activities are *interdependent* – information gained in one area may be used in another. Supervisory control tasks in power stations, for example, are concerned with doing lots of apparently simple things such as reading instruments, switching on pumps and adjusting valves. But these things must be done in accordance with decisions and judgements made earlier during other interventions and in accordance with trying to attain different specified target states. These are common features of controlling complex and potentially hazardous systems.

## INCREASE OF 'COGNITIVE' TASKS

The 1960s saw development of forms of work organisation where the operator was required to supervise semi-automated systems. Process plants, especially continuous process plants, rely on automatic control to maintain stability of important interacting parameters, such as levels, pressures and temperatures in vessels. In such plants, automation could be extensive with little manual activity required during many phases of plant operation. Manual intervention was often required during plant start-up and shutdown, including setting up processing lines, manipulating valves and sometimes, in the case of batch plant especially, handling raw materials. These entailed procedures and physical work similar to jobs that were carried out previously, but, increasingly, these manual requirements were automated. Increasing automation increased the requirement for *system supervision* with emphasis on cognitive activities such as monitoring, detection of perturbations, diagnosis, compensation for and rectification of problems and recovery. These operations were substantially *cognitive*, requiring pattern recognition, application of operating rules and principles, diagnosis and planning.

Designing and constructing large plants justified further investment in automation, measurement, control and display systems, in order to optimise productivity and reduce hazards. In nuclear power generation and petroleum refining, costs of error were so substantial that systems had to be controlled remotely for reasons of safety. In view of this demand for cognitive skill in high-risk

---

\* Of course, it could mean that they are all equally misguided.

environments, manufacturing units came to rely on display and control systems of increasing sophistication. With the increase in system supervision and decrease in manual work, the size and complexity of plants increased, prompting development of work environments that emphasised information handling and remote control.

Technology also enabled properties of materials to be measured in novel ways, information to be combined in different ways, greater amounts of information to be sampled and displayed and reliability of equipment to be improved, thereby creating conditions whereby systems were more self-regulating but increasingly challenging if self-regulation failed. Sometimes, these changes have made things simpler and sometimes they have made life more difficult for the operator. Providing digital computer-based interfaces enabled instrumentation designers to provide visual analogues of control parameters and decision aids to suit the requirements of the operators using the information. All of these changes had implications for tasks and for task analysis methods.

Developments in measurement, control and display technology have changed the characteristics of many working environments. Railway signalling and line control is increasingly automated, relying on novel forms of measurement, including the weight/occupancy of railway carriages, as well as representing traffic flows in novel and helpful ways. Air-traffic control relies increasingly on system supervision with flight-paths prescribed by computer and resolution of conflicts within airspace computed, leaving the air-traffic control officer to monitor their adequacy and intervene only as necessary. Automation and developments in robotics have prompted other industrial operations, from warehousing to vehicle assembly, to follow a similar path. As measurement methods develop, so do opportunities for further automation. Thus, many systems now require *system supervisory tasks*.

These developments created a major stimulus for human factors, including task analysis methods. Task analysis methods had to identify cognitive and non-cognitive aspects of the tasks they were addressing and try to understand how these interrelated.

The simpler methods of analysis of scientific management and work-study should not be regarded as naïve simply because they did not deal with cognition. They were useful for the sorts of work they examined and the purpose for which analysis was required. Technology has changed the landscape of work in dramatic ways, and treating work as a series of observable activities is no longer satisfactory. Task analysis methods have evolved as work practices evolved, and work practices have evolved through the development of technology.

Understanding these different challenges to operators is an important part of appreciating different demands on task analysis. An early paper by Crossman (1960) described the evolution of process plants and the attendant requirements on the human operator. Much of this paper is out of date from the perspective of modern work technology, but the account of the evolution of manufacturing work is extremely interesting and remains relevant. It explains industrial paradigms and describes their implications for what operators were then expected to do. Bainbridge (1987) discussed the ‘ironies of automation’, including many insights into how tasks changed with automation and the implications this has had for human operating requirements – less intervention is required, but such intervention is often due to unforeseen or unlikely circumstances with which the operator may be unfamiliar; hence, the operator’s ability to cope is compromised. Despite task analysis not being the main focus, each of these discussions is fundamental to task analysis, because each describes how the circumstances and signals that an operator has to deal with are created. Changes in technology result in new responsibilities for the operator, different frequencies of demand and different consequences of error – all factors that concern task analysis. Such change will, no doubt, continue into the future.

## HIERARCHICAL TASK ANALYSIS

*Hierarchical task analysis (HTA)* evolved in the 1960s and 1970s from the work of John Annett and Keith Duncan (e.g. Annett and Duncan, 1967; Annett et al., 1971; Duncan, 1972). Their work collected and organised insights and methods of other workers in the field, to demonstrate

a practical framework to guide task analysis. More recent accounts of HTA are provided by Shepherd (2001) and Stanton (2006).

Annett and Duncan were influenced by the ideas of Miller et al. (1960) who suggested cognition be represented by a hierarchy of operations and plans. *Operations* were considered units of *behaviour* undertaken in order to achieve a goal, specified in terms of a target system state. *Plans* were algorithms for organising operations. Miller et al.'s idea for decomposing an operation was to represent it in terms of a 'TOTE' unit, where TOTE was an acronym for Test-Operate-Test-Exit. The TOTE unit described how conduct of an operation entailed first making a *test* to compare the observed system state with the target state that defined the operation. If this comparison showed the system state was already compliant with the target, then no action was necessary, but an observed difference would prompt the need for an *operation* to reduce the difference.

In the earliest explanations of HTA, hierarchies of operations and plans were treated similarly to Miller et al., as behavioural elements. Then greater emphasis was placed on the functional interpretation of operations and plans. This is particularly apparent from Duncan (1972) onwards where distinction was drawn between the *categorical* descriptions of tasks in terms of their purpose and the *hypothetical* descriptions associated with speculations about constituent behaviour or the means to improve performance. This distinction is important from a practical perspective, because it enables the analyst or consultant to use task analysis as a basis for agreement with the client so that the E/HF specialist is addressing the client's requirements.

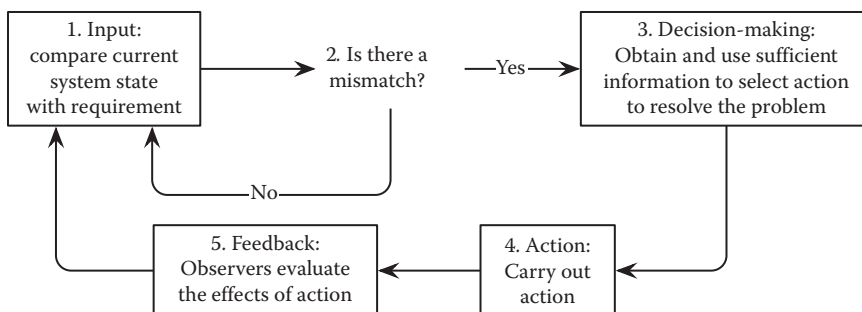
HTA describes tasks as *operations* and examines operations in terms of constituent *subordinate operations* and *plans* resulting in a hierarchy of operations and plans from which its name derives.

Annett and Duncan offered a suggestion, adapted from the TOTE unit, of using the *input-action-feedback* (I-A-F) model. Whether an *operator* is human or otherwise, an operation must entail:

- An *input* – of information in order to – prompt action and inform judgement about whether anything needs to be done to move the system being controlled towards the target state, and decision-making in order to select a course of action
- An *action* – selected as appropriate – to address deviation from the target state
- Further information to provide *feedback* to confirm or otherwise that the action was successful, indicate whether anything further needs to be considered and to regulate an on-going action.

This model represents a form of cognition, but not necessarily human cognition. Rather, it considers the minimum that would be necessary for any agent to fulfil an operation, i.e. it is an operational description.

I-A-F is devoid of an explicit stage that entails decision-making, although decision-making can be inferred. This can be made explicit by introducing a cognition 'decision-making' stage, as in Figure 6.4, without violating the principle that an operation is neutral with respect to psychology.



**FIGURE 6.4** The processes implicit within an operation – an extension of I-A-F.



At stage 1, information is obtained to compare the observed with the target state. Satisfaction means that nothing need be done other than returning to monitoring future deviations. Dissatisfaction prompts decision-making [3] to select action [4]. The selected action must be monitored [5] in order to regulate it and decide whether and when it has achieved its purpose.

The explanation of I-A-F, and its expansion in Figure 6.4, distinguishes between a *goal/objective* and its associated *operation*. Any operation has a goal/objective to which it is directed, and each objective implies an operational sequence that will achieve it.

Annett and Duncan's work began as a project to address task analysis method in general and to account for both cognition and action within the single analysis project. Rather than proffering a specific technique, it was intended as a *framework* for undertaking task analysis projects to serve the requirements of effective applied psychology and ergonomics. Their earlier papers referred to 'task analysis'. The word 'hierarchical' was added later to distinguish the approach. In some respects, this was unfortunate because it then made *HTA* a commodity, to be compared alongside similar commodities. It might have been better had the idea of a *framework* persisted, because then, it would have enabled debate regarding how different stages might be accomplished rather than judging whether the overall method should stand or fall.

## COGNITIVE TASK ANALYSIS

With the application of computers in the workplace and the consequent changes of types of work emphasising cognition, several writers have argued the need for specialist CTA methods. Finding a general definition of CTA is as difficult as finding a general definition of task analysis. One useful definition is that CTA techniques seek to model the mental activity of a task operator (May and Barnard, 2004). Because cognition is so critical, the idea of CTA is appealing with many would-be analysts judging, at the outset of their project, that a CTA approach must be adopted. Some analysts do a bit of each, using HTA to situate their work within a manufacturing or commercial situation and as a basis for discovering where cognition is critical and warrants more rigorous attention.

CTA became prominent in the 1970s and 1980s, but many previous investigators had also been concerned with the analysis of cognition in tasks. This was evident from the work of Crossman, Miller, Annett and Duncan and many others, such as Gagné and Fleischman. Gagné (e.g. Gagné et al., 1988) presented a hierarchical model of learning where simple cognitive operations were subsumed by more complex layers of rules and principles. Evidence of these relationships could be demonstrated through transfer of training studies. Fleischman and Quaintance (1984) offered an approach to the structure of skills, where complex skills depended on the mastery and execution of simpler component skills. Their work was derived from factor analytic studies. In this way, Gagné, Fleischman and colleagues presented a type of cognitive architecture before ideas of cognitive architecture became prominent through work such as Card et al.'s GOMS (1983), Anderson's ACT\* (1990) and Johnson et al.'s TAKD (1984). These later approaches aimed to demonstrate how task behaviour emerges from fundamental knowledge structures and cognitive processes, often using analogies from computer architecture and computer programming languages.

A well-known cognitive approach is GOMS (Card et al., 1983). GOMS is an acronym describing, briefly:

- *Goals* – what an operator or user seeks to accomplish.
- *Operators* – basic perceptual, motor or cognitive actions used to accomplish the goals. These are regarded as primitives, the building blocks of complex activity.
- *Methods* – procedures describing suitable ways to accomplish goals in terms of subgoals and operators. Different methods may be used to achieve specific goals, reflecting the level of expertise of the user/operator or the rigour with which a procedure must be carried out to minimise risk for example.
- *Selection rules* – specifying which method should be used on a given occasion.



An application of GOMS was the keystroke-level model, used to compare computer dialogues by counting keystrokes needed to accomplish various operations in, for example, a word processor application. Thus, different computer applications could be compared. GOMS is presented as a cognitive approach, but it uses much of the language of existing task analysis methods. *Goal* is used to describe the user's purpose. *Operators* are behavioural primitives. They are different to *operations* in HTA, because operators refer to standard units of action, much like what Fleischman and Quintaince developed, whereas operations, in HTA, for example, are any process that achieves a goal. *Methods* are like *plans* in HTA, in particular procedural plans. *Selection rules* relate to the contingencies encountered in variable sequence plans in HTA.

ACT\*, an acronym for 'adaptive character of thought', is a model of cognition designed to explain how declarative knowledge is compiled to respond to more complex or novel circumstances. Declarative knowledge is expressed in the form of production rules ('if – then' statements). An operator may compile several such production rules, in the manner of programming languages such as *LISP* to generate a new 'procedural' rule. In the case of fault-diagnosis, for example, such a process may help an operator infer a relationship between variables that is not evident simply by considering the original production rules in isolation.

While GOMS and ACT\* focus on the structure of cognition, other approaches focus on the processes of cognition. Prominent are Reason's GEMS (generic error modelling system; Reason, 1990) and Rasmussen's SRK model (skills–rules–knowledge–Rasmussen, 1980). Both of these acknowledge that human operators with different levels of expertise and confronting problems of different degrees of familiarity may respond in different ways. For example highly consistent events may be dealt with by someone with suitable familiarity in a very automated skill-based manner. Less familiarity, or less confidence in offering a stereotyped response, may result in using rules to generalise from other known responses or resort to reasoning from basic principles. The interaction between expertise and the response method followed will, it is argued, result in different patterns of error. A task analyst using one of these taxonomies makes a judgement about the type of behaviour deployed in a response and uses this to predict error type or to propose a design remedy.

One practical concern with trying to account for cognition is the danger of assuming all people undertaking the same work activity organise cognition and action in the same way. While the laws governing cognitive processes may apply between people, it is unlikely that the same behavioural structure and content holds for different people. Cognition is private and different people adopt different strategies according to their previous experience. They vary in their personal capabilities, style and preferences in handling information. The privacy of cognition means that it is impractical to validate any of these representations anyway. Basing a task analysis on someone who has demonstrated they are a *competent* operator is hardly satisfactory, because competency is judged in terms of *performance* and different private processes can lead to equivalent public performance. Moreover, people may be at different stages of development. Several researchers have argued for the compilation of knowledge and skill from simpler components through the medium of practice (e.g. Fitts, 1965; Rasmussen, 1980; Anderson, 1990).

## TASK CLASSIFICATION AND TABULAR TASK ANALYSIS

A common practice in task analysis projects is the use of classification or categorisation schemes. The rationale for any such scheme is that by establishing a task element as a member of a category, it can be treated in the same way as other members of that category. It is common to encounter this sort of approach being used in human error analysis, for example SHERPA (systematic human error reduction and prediction approach – Embrey, 1986). SHERPA is a methodology that entails conducting a task analysis using HTA, then assigning the actions assumed within the operation to a standard set:

- Action (e.g. pressing a button or pulling a switch)
- Retrieval (e.g. getting information from a screen or manual)

- Checking (e.g. conducting a procedural check)
- Selection (e.g. choosing one alternative over another)
- Information communication (e.g. talking to another party)

Following this, and the application of keywords, the analyst looks for potential for error within the task in accordance with error modes for the particular type of operation. Such approaches depend upon the classification scheme being exhaustive; otherwise, the analyst will shoehorn an operation into a category and this might be inappropriate. Further illustration of this approach can be seen in the application of SRK or GEMS.

A concern with approaches such as SHERPA is that it may fail to take account of the operating context in which the instance of the particular activity resides – in one context, the operator may be primed for action in a way that is impossible in another context. There is also a risk in assuming that apparently similar operations imply that they will always depend upon the same behaviour to carry them out, irrespective of who is responsible. There is a danger if the analyst thinks the task analysis process is now easy simply because there is a *checklist* to be applied that will make everything correct. Even when using a classification scheme, the analyst needs to scrutinise task context to judge whether an instance of a particular operation is genuinely representative of a category.

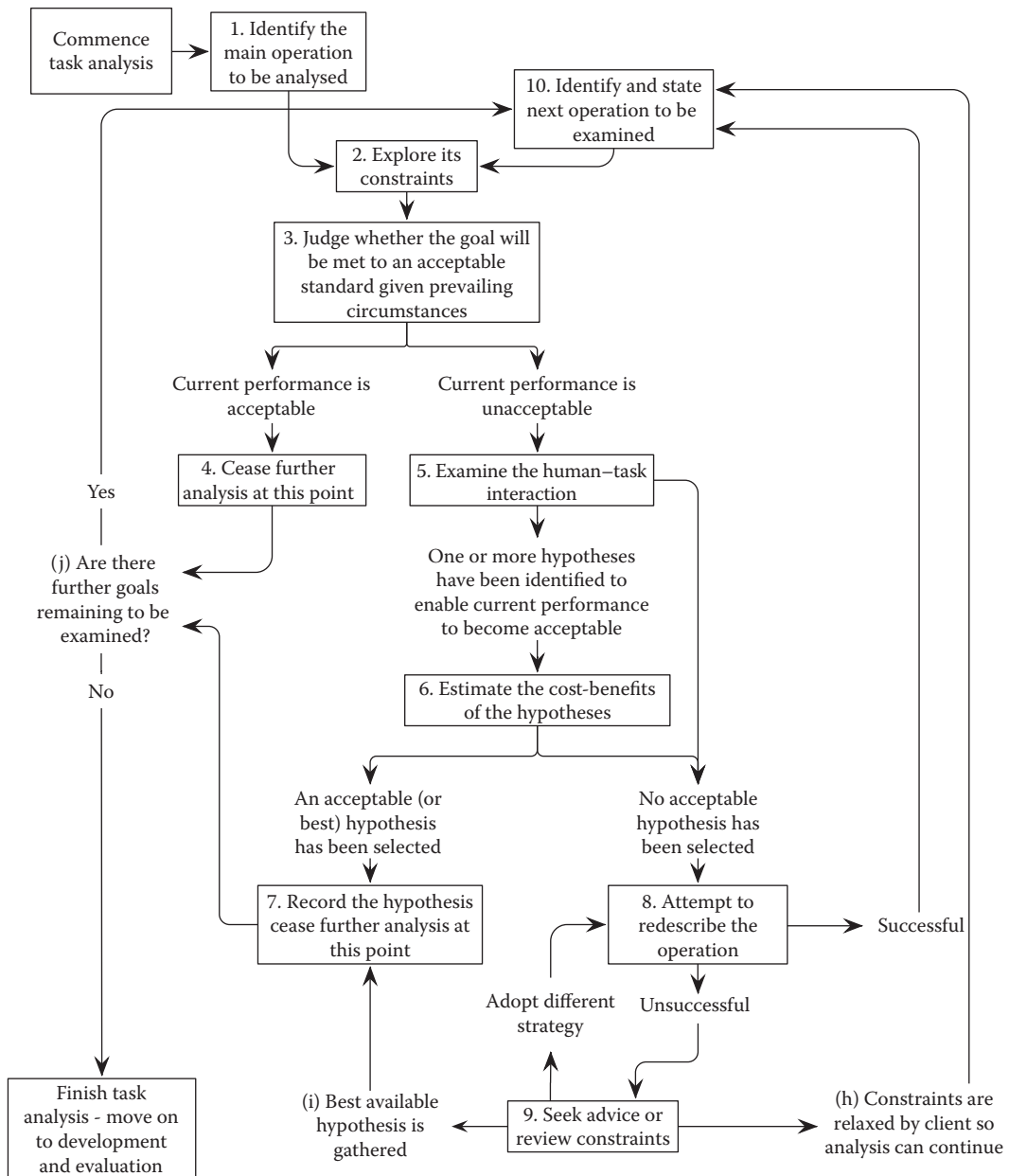
## HTA AS A TASK ANALYSIS FRAMEWORK

It was emphasised that HTA was initially developed as a framework for task analysis. It was an invitation to consider how best to make sense of complex tasks. This section will revisit HTA as a task analysis framework. It suggests a distinct way of progressing task analysis that still enables other techniques to inform the decisions that have to be taken on the journey to completion. It offers a way of structuring a task description which enables appropriate complexity to be dealt with and the cohesion of the overall task to be checked. It also provides a systematic routine for identifying hypotheses aimed at dealing with the various task problems identified.

The main features of HTA are:

|                           |                                                                                                                                                                                                                                                                                                                  |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Operational description   | Tasks should be analysed by describing performance of people in operational terms, i.e. what has to be achieved rather than behaviour employed in meeting targets.                                                                                                                                               |
| Hierarchies of operations | Operations can be organised hierarchically. This allows the ‘grain’ of analysis to be adjusted. It enables operations to be described in sufficient detail to enable task influences to be best considered by the analyst. The hierarchical organisation does not imply the organisation of human cognition.     |
| Plans                     | Subordinate operations are governed by ‘plans’ that specify conditions when each subordinate operation should be carried out to satisfy the requirements of their common superordinate operation. Plans are a requirement to complete redescription and not intended as a component of human cognition.          |
| Redescription             | Task analysis proceeds by redescribing operations as a set of subordinate operations and a plan specifying the conditions when subordinate operations are carried out to enable their common superordinate operation.                                                                                            |
| Hypotheses                | Following the <i>categorical</i> description of performance in operational terms as aforementioned, the analyst judges the effectiveness with which a particular operation is carried out, with a view to identifying weaknesses or proposing ways in which the operation would be carried out more effectively. |
| Stopping rules            | Rational <i>stopping rules</i> are declared and employed by the analyst in order to prevent the process of redescription being continued unnecessarily.                                                                                                                                                          |

Implicit in these six concepts is a flow of decisions that analysts will take *en route* to completion of the task analysis. This flow of decisions is captured in the diagram in Figure 6.5.



**FIGURE 6.5** The flow of decisions implicit in the HTA process, providing a framework for task analysis.

HTA focuses on a main work goal [1]. The analyst needs to consider this goal and the constraints that influence performance [2]. This is most sensibly done by using a task expert as an informant. Together, the analyst and task expert judge whether the goal will be met to an acceptable standard [3] and the outcome of this decision causes a different route to be followed. The main challenges arise where current performance is judged unacceptable. In these cases, the analyst must consider the human-task\* interaction [5] and seeks a hypothesis concerning a cause of this problem or a solution

\* The term 'human-task interaction' is preferred to 'human-machine' interaction in this context, because it is less likely to constrain thinking about the factors that influence performance.

to the problem. This depends upon the analyst's knowledge or the use of additional methods such as task taxonomies or CTA, for example. Often, the analyst has experience of how such things might be done better, or the client or task informants make suggestions. Another way is to research the issue. An implicit cost-benefit analysis [6] enables the value of any such hypothesis to be assessed. Sometimes, it is difficult to work out how to complete a redescription, so advice is sought either concerning strategy to analyse further or insights concerning appropriate hypotheses [9]. When these processes have been completed, the result is the hierarchy of operations and plans, with their attendant hypotheses, which characterise HTA.

## TYPES OF PLAN

A key feature in organising a task hierarchy, certainly according to the HTA framework, is the representation of plans. There are many ways in which sub-operations can be carried out to attain different goals and, for most practical purposes, there is little to be gained by standardising them. It is useful, however, to recognise some of the ways in which plans can vary.

### Fixed Sequences

The simplest type of plan is a *fixed* sequence, where a specified second operation is carried out when a first operation has been successfully attained. This means successful feedback to a first operation is obtained – this is the cue to carry out the second operation, and so on. This sort of plan is encountered in straightforward tasks such as some computing applications. Saving text in Microsoft Word, for example, can be achieved by the following:

1. Move cursor to the *File* menu.
2. Hold down mouse-button over the word *File*.
3. Move cursor down to the menu item *Save*.
4. Release mouse-button.

This will result in the document being saved.

*Do 1 – 2 – 3 – 4* is a *fixed sequence* plan, because each action is done in order when it is apparent that the previous operation in the sequence has been completed. It is important to note that, logically, fixed sequences are not simply strings of actions but operations that each involves monitoring feedback to ensure the actions were successful.

### Cued Actions

Another relatively simple but distinct plan is a cued action. The operator may have several things that might be done but has to wait for an authorised person to say what will be done on this occasion. Logically, a cued action plan would require the following set of operations:

1. Monitor instruction.
2. Carry out operation A.
3. Carry out operation B.
4. Carry out operation C.

And the plan would state: *Do 1. Then according to instructions do 2, 3 or 4.*

The importance of stating a *cued action* plan is that it prompts the analyst to recognise that a first operation must be included in order to consider why the operator should choose one of operations A, B or C. Sometimes this is a verbal instruction from a colleague; sometimes, it will be the response to a signal such as an alarm. It is rule-based.

### Contingent Fixed Sequence

A *contingent fixed sequence* plan is where operations are carried out in a predetermined order, but the cue to act for some of these operations might be system feedback to demonstrate the system has changed as a consequence of a previous operation. This differs from the *fixed sequence* plan where the queue for the next operation is feedback showing the previous operation has been completed.

Plan elements such as this occur in tasks concerned with controlling large systems, including engineered systems, socio-technical systems and large organisations – a first operation may have caused a change to operating conditions that then caused system parameters to change, process control situations being obvious examples.

One important class of cues is the instruction from a manager, supervisor or other colleague. This is particularly important in complex systems where several people must collaborate for a successful outcome. Often a person knows what to do next, but must wait for someone else, with authority, to say when it should be done. Authority is used here in a broad sense and not simply in the sense of managerial control. A salesperson may need to wait for the word from a production supervisor to say when fresh a product will be available before clients can be told when they may expect it. Other examples can be given from hazardous environments, where one operator overseeing one crucial activity may provide the cues for others to carry out operations to ensure that everything comes together at the right time. Coordinated actions or communications within teams can often be accounted for by cued actions. Part of a railway-signalling task might be *set routes for train to progress*. This could be redescribed as follows:

1. Establish whether train is permitted to progress.
2. Set route as far as possible: push signal buttons on display.
3. Establish that route has been set: review appearance of white lights on track display.

The contingent fixed sequence plan might say: *Do 1. If train can progress – 2 – 3.*

Plans such as these are commonplace and important. It is important that analysts take care to avoid oversimplifying the task by opting for a fixed sequence plan when a little more attention might reveal that contingencies need to be observed.

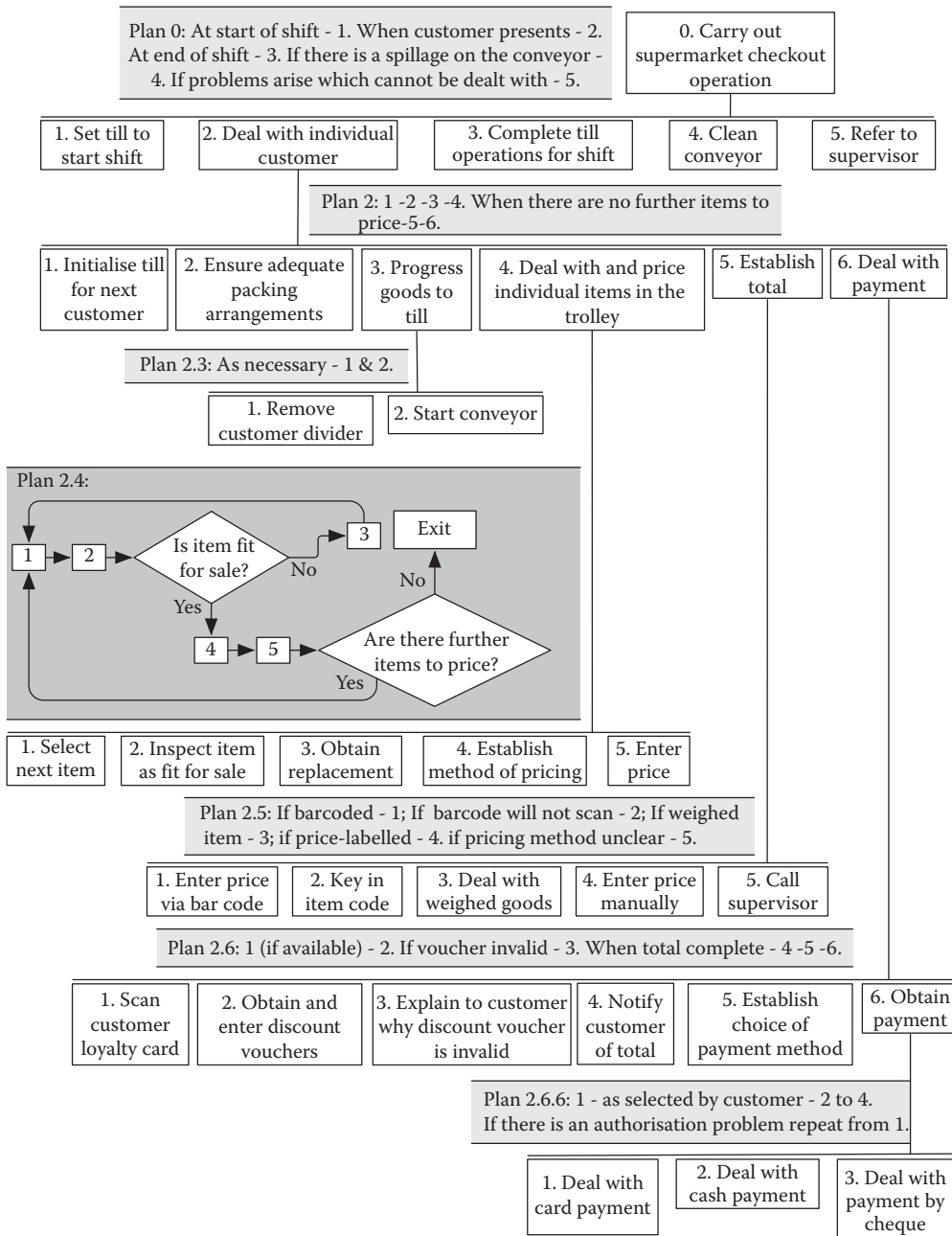
### Choices

Choices are logically similar to cued actions in that one of a set of subordinate operations is carried out according to prevailing conditions. This means particular attention needs to be paid to an initial decision-making operation. When choice is offered, it may not be the alternatives available that will create the problems for an operator but rather the decision concerning which one to choose. Sometimes making a choice can be left to the discretion of the operator. More often there is a reason for doing one thing rather than one of several others. The obvious example is fault diagnosis. Choices are particularly important, because they prompt the analyst to consider the judgement and skill necessary in order to make the appropriate choice. These are not simple rule-based choices, because the choice made depends upon the reasoning of the operator. Following this reasoning, selecting the next action could be said to be rule-based.

### Cycles

In some tasks, people must repeat a sequence of operations until conditions arise when they must stop. A cycle is a combination of fixed sequences and a cue to indicate when stopping is required. Sometimes the cue to stop is given by another person, for example a supervisor. This is typically the case in repetitive work situations. Sometimes the cue to stop is provided by a test result. This is typical in situations where targets must be attained and action sequences must continue until this criterion is reached.

Cycle plans are very common especially in customer service tasks including telephone sales and operating supermarket checkouts. Generally, they apply to dealing with clients or customers. Often, cycles contain operations which are themselves cycles. For example a supermarket checkout operator will conduct a cycle commencing with each new customer, then conduct further cycles as each item is picked from the conveyor belt, processed and priced, then moved to the packing area (see Figure 6.6).



**FIGURE 6.6** Example of HTA as applied to a supermarket checkout task. This example contains a variety of plan types that are characteristic of many tasks in industry and commerce.

### Time-Sharing

Sometimes two (or more) operations have to be carried out together. This must be stated explicitly in the plan. Time-sharing elements are very common. Sometimes time-sharing creates serious problems for the person carrying out the operation in terms of task loading and demands on attention. Sometimes time-sharing places particular stress on the operator having to attend to two things simultaneously. In other cases, it is possible to choose to devote attention to one subordinate operation until it can be safely put on hold, then switch attention to the other. For example a swimming pool attendant or a nurse supervising a ward must continually scan the people in their charge in order to establish their well-being. Recognising a problem can capture attention, causing neglect to others who may then find themselves in trouble. Knowing how to do this reflects growing expertise, but it may also depend upon people having the right personality to maintain a balance between attention in breadth and focus.

### COMBINING PLAN ELEMENTS AND PLANS

The plan elements in the example have been outlined to show features of plans that might be encountered. It is not important that plans conform to one or other of these types, but it can be helpful to know that these are the common features encountered. In practice, actual plans may contain several such features which combine to specify the conditions necessary for certain subordinate operations to satisfy their common superordinate operation. In this way, more complex task structures can be achieved.

Sometimes plans appear too complex to understand and state clearly. In these cases, it is often possible to identify subgroups of a set of subordinate operations where each member of the subgroup shares a common goal. This allows the analyst to inject a new 'operation' reflecting that common goal that governs a few subordinate operations for which it is easier to write a plan. Equally the new operation replaces several operations that the analyst had to try to understand in the original configuration. This is a neat technique for enabling a full range of plans to be expressed instead of a cluttered mess. Thus, one very complex plan that might have been too difficult to express has been replaced by a hierarchy of far clearer plans and intermediate operations.

### REDESCRIPTION RATHER THAN DECOMPOSITION

When an operation is redescribed, it is important to do this as explicitly as possible by stating plans fully. Doing this can draw attention to operations that may be overlooked from analysis, especially cognitive operations responsible for decision-making. Cases in point are the various monitoring and decision stages in *cued actions*, *contingent fixed sequences*, *choices* and *cycles* discussed previously. It is easy for task experts to overlook the crucial cognitive aspects of their skill and concentrate on observable action. It is also tempting for a task analyst simply to follow this lead and overlook important elements of cognition, rather than using the plans to probe further.

It is inappropriate simply to treat the hierarchical structure as a convenient way of decomposing the task in order to identify constituent elements to which to apply, say, a task taxonomy. A full redescription with a proper plan often identifies operations and decisions that might have been overlooked.

### STOPPING RULES

Annett et al. (1971) presented the 'PxC rule' as the main stopping rule. Here, the analyst estimates the product of *probability of inadequate performance* and *cost of inadequate performance* – this is a rule-of-thumb, rather than a calculation, although a calculation would be possible. It is useful in justifying further effort in redesign or other change. It is a form of risk analysis. It is not the only stopping rule that could be used. In interface design, a stopping rule could be expressed in terms of



whether the operation implies *action* or *cognition* that deals primarily with information or control at the interface. Where a categorisation scheme is used to consider each operation in turn, such as in human reliability assessment (see Chapter 30), it may be judged appropriate to stop when an operation is encountered that conforms to a category within a classification scheme.

### AN ILLUSTRATION

It is helpful to illustrate this with an example. Providing illustrations of task analysis methods is difficult, because readers may be unfamiliar with the work being described in a particular example. It can be beneficial to describe more familiar everyday tasks, but these may be regarded as unconvincing. A good one, though, is that of a supermarket checkout operator (illustrated in Figure 6.6). This task is complex in a way that is reflected in the complexity of many real work tasks. It will vary from store to store and between generations of technology. It has changed over the past 20 years in view of varying sales methods and varying processing technologies, but it still remains a valid example to observe in order to understand how this and, indeed, many other work tasks are structured.

The top-level plan in Figure 6.6 concerns engaging with a work task and then withdrawing when the shift is over. The example provides a number of complex plans that are quite typical of many jobs. Another common feature that makes this task so representative of many work tasks is the presence of *cycles* of activity, where sequences of operations are repeated until an end point is reached. This is common in system supervisory tasks, sales tasks, maintenance tasks, clerical tasks, management and supervision. These plans are particularly interesting in describing the apparent complexity of work, as cycles are undertaken within the context of higher order cycles.

### OPERATIONS, GOALS AND THE RE-SCOPING OF LEVELS

A problem in any task analysis is knowing how broadly or how narrowly to view the tasks to be analysed. Operating problems often come to attention by virtue of a pattern of errors associated with particular operations. There is the temptation to focus attention on that operation and not elsewhere. The operator of a batch chemical plant might be criticised, because he or she is taking too long to adjust a mixture to the correct formulation. It may be assumed that the operator is insufficiently skilled and needs training. Concentrating on this specific operation might prompt ways to improve instrumentation or training. However, analysing the task at a more general level could show that the formulation problem was due to insufficient cleaning and inspection of vessels during plant preparation following a previous batch and this is the problem that really needs to be solved.

In task analysis, there is often benefit in expanding the scope of the analysis, because operations are interdependent and a solution to a problem elsewhere might be simpler and eradicate the more difficult problem to solve that was initially addressed. As a general rule, proper preparation, including cleaning equipment, setting up equipment, adjusting settings and ensuring adequate resources are available, is critical to performance requirements later on. In a similar way, operator engagement in earlier phases of the task provides information and expectations that will generally serve later activities where this information is required. Widening scope also clarifies how present operations must be done in order that conditions are set most favourably for operations yet to be done.

All of these issues can be addressed by widening the scope of the task being examined. It is quite likely that such revision is indicated at a later stage in the analysis, in which case re-scoping would need to be done then.

### INHERITANCE OF CONTEXT

A benefit of hierarchical description concerns inheritance of context. If an overall operation is subject to various risks, then its subordinate operations may also be subject to the same risks. If some of its subordinate operations are not subject to the risk, then their subordinate operations will not

inherit the risk either. This is a convenient way for focusing on areas of concern and also recognising where the operator is subject to stress.

An important feature of inheritance is where a plan higher in the hierarchy specifies *time-sharing*. It means that any subordinate operation is now carried out in a context where it may need to be carried out with other operations. Behaviour appropriate to an operation carried out on its own may no longer be appropriate in these time-shared circumstances.

Generally, the hierarchy specifies where any particular operation resides within the wider pattern of activity. This can change how different operations may be treated.

## BRINGING THE TASK INTO FOCUS

The final point, which is fundamental to analysis, is that through a rational approach of redescription, various aspects of the task are systematically brought into focus. While it may be acknowledged that there are lots of influences from the system on the behaviour of the operator, enabling this focus helps the analyst recognise which features of the task influence which aspects of performance. For example, infrequency and inconsistency will mean that the operator cannot rely on a stereotyped response to related circumstances. Equally, some physical work that has to be undertaken in difficult areas of a plant will be more susceptible to extreme environmental conditions than cognitive activity carried out in a control room where the environment is not extreme.

## EXAMINING BEHAVIOUR APPROPRIATE TO OPERATIONS

Identifying an operation does not mean that a distinct set of constituent behaviours can be described. Operations are neutral with respect to solution. Operations do not have to be carried out by a particular method, especially where alternative technologies are available to furnish a solution. Automation, for example, is a solution that may leave the operator with little to do. Reassigning an operation from one person to a colleague who is already competent is a satisfactory solution. Improved interface design might enable people to perform more effectively with minimal instruction. Job-aids or use of a local expert might enable an operation to be carried out satisfactorily without radical redesign and with minimal training. The supermarket checkout task provides clear evidence of the benefits of observing this neutrality. Over the past 25 years, we have witnessed the shift in 'deal with and price individual items...' as moving from operators reading price labels, through the use of bar codes, through customer self-service and even operators memorising item costs. Technology and customer-service considerations have changed, but the purpose stays the same.

Training would feature as an option in supporting most operations dependent upon human intervention, but training needs to be adaptive to accommodate individual differences. All of these options offer considerable flexibility for task analysis and human factors designers. An operation does not imply a distinct pattern of behaviour applied to each occasion that people carry out that operation.

Human behaviour with regard to the execution of tasks is generally described in terms of psychological processes. Such processes are *constructs* that psychologists have used in seeking to understand and explain behaviour. These constructs include attention, memory, decision-making, reasoning, inferring and so on. Miller's model in Figure 6.3 is such an approach.

## VALIDATING A TASK ANALYSIS

With any E/HF method, there are questions of reproducibility and validity. Is it important that two analysts would arrive at the same outcome in terms of what was recorded about the task? The answer is probably 'no' in most cases. Generally, task analysis is treated as a tool to aid design and evaluation of performance in systems. Differences in how a task is described relate to the different experiences of the respective analysts and their knowledge of how the wider system impinges on

behaviour and how behaviour impinges on task demands. Task analysis is a tool in the sense that it helps us do things better than we would have done without the tool. It is not a method of scientific measurement, where observational consistency would be demanded. Task analysis is beneficial if it helps answer the *next* questions in system design or evaluation in a way that leads to a better outcome. The better outcomes sought when task analysis is used consist of improved designs for workplaces, documentation, software, personnel selection and training, and it also leads to better insights into areas such as error analysis which result in greater assurance that systems are safe and effective. Even if there are no practicable ways of measuring the reliability and validity of task analysis methods themselves, their benefits should still be judged in terms of the effectiveness of the decisions they inform and the developments that derive from these.

## CONCLUSIONS

There has been a great deal of effort expended in developing task analysis methods and methodologies, but comparatively little attention to resolving what tasks and task analysis are really about or the extent to which different approaches are similar or different. The field of task analysis has tended to be one that packages specific methods developed in a particular type of domain to provide a process to follow in analysing tasks, then attaching an acronym for the purposes of *branding*. There has been little effort in reviewing and generalising basic principles. One exception was the work of Annett and Duncan (1967) that identified ideas and strategies that should be taken into account when analysing tasks. Unfortunately, their task analysis framework, which comprised ideas and principles derived from other researchers, was itself branded as HTA. It then became regarded as a self-contained method that stood or fell in its entirety. A more recent attempt has been provided by Diaper and Stanton (2004) in the final chapters of their handbook. They revisit basic ideas and the authors suggest creating a unifying scheme. But even if such a unifying scheme were devised, it is difficult to believe that it would be wholeheartedly adopted in practice, when so many people have invested in their own approaches. It would be a massive enterprise and would have to solve an immense number of difficulties. It may not be worth the effort.

One major conflict is the contrast between pragmatic approaches, where task analysis methods focus on the requirements to support systems at work, and approaches which are concerned to provide a defensible model of human behaviour if they are to be regarded as valid. Most people undertaking task analysis in industrial and commercial situations do so from the perspective of understanding impediments to performance and generating solutions for improvement, irrespective of whether this is accompanied by an accurate model of cognition.

In examining principles of task analysis, the chapter has considered the principles expressed by Annett and Duncan that were simply presented to encourage good practice. Had Annett and Duncan's ideas continued to be regarded more as a set of principles than as a specific brand, they would have contributed to a more effective debate of how best to analyse tasks, including incorporating more ideas for general benefit.

## KEITH DUNCAN: AN APPRECIATION

It is with sadness that the publication of this chapter on task analysis closely follows the death of Keith Duncan in December, 2013. Keith, who together with John Annett, gathered the key principles for effective task analysis, then organised and presented them as a coherent framework that became known as hierarchical task analysis (HTA). HTA has now been developed, used, abused, misunderstood and misrepresented for nearly 50 years. Keith's approach was always highly principled and pragmatic. HTA has made, and continues to make, an invaluable contribution to applying psychology and human factors, ensuring that outcomes of applied work address real needs. On a personal level, Keith was an excellent mentor and colleague to whom many people, such as myself who were privileged to work with him, owe an enormous debt.

## REFERENCES

- Anderson, J.R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Annett, J. and Duncan, K.D. (1967). Task analysis and training design. *Occupational Psychology*, 41, 211–221.
- Annett, J., Duncan, K.D., Stammers, R.B. and Gray, M.J. (1971). *Task Analysis*. London, U.K.: Her Majesty's Stationery Office.
- Bainbridge, L. (1987). Ironies of automation. In Rasmussen, J., Duncan, K.D. and Leplat, J. (Eds.), *New Technology and Human Error*. Chichester, U.K.: John Wiley & Sons.
- Card, S., Moran, T.P. and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crossman, E.R.F. (1956). Perceptual activities in manual work. *Research*, 9, 42–49.
- Crossman, E.R.F. (1960). *Automation and Skill*. London, U.K.: H.M.S.O.
- Currie, R.M. (1977). *Work Study*. London, U.K.: Prentice Hall.
- Diaper, D. (1989). *Task Analysis for Human-Computer Interaction*. Chichester, U.K.: John Wiley & Sons.
- Diaper, D. and Stanton, N. (2004). *Handbook of Task Analysis for Human Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Duncan, K.D. (1972). Strategies for the analysis of the task. In Hartley, J. (Ed.), *Programmed Instruction: An Educational Technology*. London, U.K.: Butterworth, pp. 19–81.
- Duncan, K.D. (1974). Analytical techniques in training design. In Edwards, E. and Lees, F.P. (Eds.), *The Human Operator and Process Control*. London, U.K.: Taylor & Francis Group.
- Embrey, D.E. (1986). SHERPA: A systematic human error reduction and prediction approach. Paper presented at the *International Meeting on Advances in Nuclear Power Systems*, Knoxville, TN.
- Fitts, P.M. (1965). Factors in complex skill training. In Glaser, R. (Ed.), *Training Research and Education*. New York: Wiley, pp. 177–197.
- Fleischmann, E.A. and Quintaince, M.K. (1984) *Taxonomies of Human Performance*. New York: Academic Press.
- Gagné, R.M., Briggs, L.J. and Wager, W.W. (1988). *Principles of Instructional Design*. New York: Harcourt Brace Jovanovich.
- Gilbreth, F.B. (1912). *Primer of Scientific Management*. New York: Van Nostrand.
- Johnson, P., Diaper, D. and Long, J. (1984). Tasks, skills and knowledge: Task analysis for knowledge-based descriptions. In Shackel, B. (Ed.), *INTERACT'84 – Proceedings of the First IFIP Conference on Human-Computer Interaction*. Amsterdam, the Netherlands: North Holland.
- Kirwan, B. (2005). Human reliability assessment. In Wilson, J.R. and Corlett, E.N. (Eds.), *Evaluation of Human Work*, 3rd edn. London, U.K.: Taylor & Francis Group.
- Kirwan, B. and Ainsworth, L.K. (1992a). *A Guide to Task Analysis*. London, U.K.: Taylor & Francis Group.
- Kirwan, B. and Ainsworth, L.K. (1992b). Appendix: Summary of task analysis techniques. In Kirwan, B. and Ainsworth, L.K. (Eds.), *A Guide to Task Analysis*. London, U.K.: Taylor & Francis Group, pp. 391–402.
- May, J. and Barnard, P.J. (2004). Cognitive Task Analysis in Interacting Cognitive Subsystems. In Diaper, D. and Stanton, N. (Eds.), *Handbook of Task Analysis for Human Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Miller, G.A., Galanter, E. and Pribram, K.H. (1960). *Plans and the Structure of Behavior*. New York: Holt, Reinhart & Winston.
- Miller, R.B. (1962). Task description and analysis. In Gagné, R.M. (Ed.), *Psychological Principles in Systems Development*. New York: Wiley, pp. 187–228.
- Miller, R.B. (1966). Task taxonomy: Science or technology. In Singleton, T., Easterby, R.S. and Whitfield, D.C. (Eds.), *The Human Operator in Complex Systems*. London, U.K.: Taylor & Francis Group.
- Miller, R.B. (1967). Task taxonomy: Science or technology. In Singleton, W.T., Easterby, R.S. and Whitfield, D.C. (Eds.), *The Human Operator in Complex Systems*. London, U.K.: Taylor & Francis Group, pp. 187–228.
- Rasmussen, J. (1980). The human as a systems component. In Smith, H. and Green, T.R.G. (Eds.), *Human Interaction with Computers*. London, U.K.: Academic Press, pp. 67–96.
- Reason, J.T. (1990). *Human Error*. Cambridge, U.K.: University Press.
- Seymour, W.D. (1966). *Industrial Training for Manual Operations*. London, U.K.: Pitman.
- Shepherd, A. (2001). *Hierarchical Task Analysis*. London, U.K.: Taylor & Francis Group.
- Stammers, R.B. and Shepherd, A. (2005). Task Analysis. In Wilson, J.R. and Corlett, E.N. (Eds.), *Evaluation of Human Work*, 3rd edn. London, U.K.: Taylor & Francis Group.
- Stanton, N.A. (2006). Hierarchical task analysis: Development, applications, and extensions. *Applied Ergonomics*, 37, 55–79.
- Wickens, C.D., Hollands, J.G., Banbury, S. and Parasuraman, R. (2013). *Engineering Psychology and Human Performance*, 4th edn. New York: Harper Collins.

---

# 7 Knowledge Elicitation

## *Methods, Tools and Techniques*

*Nigel R. Shadbolt and Paul R. Smart*

### CONTENTS

|                                                                          |     |
|--------------------------------------------------------------------------|-----|
| Introduction.....                                                        | 163 |
| Knowledge Elicitation Techniques .....                                   | 166 |
| Interviews .....                                                         | 167 |
| Protocol Analysis .....                                                  | 170 |
| Critical Decision Method .....                                           | 174 |
| Concept Sorting.....                                                     | 176 |
| Repertory Grids.....                                                     | 178 |
| Laddered Grids.....                                                      | 180 |
| Limited-Information Task .....                                           | 183 |
| Concept Mapping and Process Mapping.....                                 | 183 |
| Classification of Knowledge Elicitation Techniques.....                  | 184 |
| Experts and Expertise .....                                              | 186 |
| Methodologies and Programmes.....                                        | 188 |
| Knowledge Elicitation Tools.....                                         | 191 |
| PCPACK.....                                                              | 192 |
| Protégé.....                                                             | 192 |
| CmapTools .....                                                          | 193 |
| Knowledge Elicitation, Knowledge Engineering and the World Wide Web..... | 194 |
| Conclusion .....                                                         | 196 |
| References.....                                                          | 196 |

### INTRODUCTION

Knowledge elicitation consists of a set of techniques and methods that attempt to elicit the knowledge of a domain expert,\* typically through some form of direct interaction with the expert. Knowledge elicitation is a sub-process of knowledge acquisition (which deals with the acquisition or capture of knowledge from any source), and knowledge acquisition is, in turn, a sub-process of knowledge engineering (which is a discipline that has evolved to support the whole process of specifying, developing and deploying knowledge-based systems).

Although the elicitation, representation and transmission of knowledge can be considered a fundamental human activity – one that has arguably shaped the entire course of human cognitive and social evolution (Gaines, 2013) – knowledge elicitation had its formal beginnings in the early to

---

\* It should be pointed out that although early conceptualisations of knowledge elicitation cast the process as one of extracting or mining knowledge from the heads of experts, more recent conceptualisations view the process as a modelling exercise. The idea is that the knowledge elicitor and domain expert work together in order to create a model of an expert's knowledge. This model may reflect reality to a greater or lesser extent.



mid-1980s in the context of knowledge engineering for expert systems.\* These systems aimed to emulate the performance of experts within narrowly specified domains of interest,<sup>†</sup> and it initially seemed that the design of such systems would draw its inspiration from the broader programme of research into artificial intelligence. In the early days of artificial intelligence, much of the research effort was based around the discovery of general principles of intelligent behaviour. Newell and Simon's (1963) General Problem Solver exemplified this approach. They were interested in uncovering a general problem-solving strategy that could be used for any human task. In the early 1970s, however, a new slogan came to prominence: 'in the knowledge lies the power'. A leading exponent of this view was Edward Feigenbaum from the Stanford Research Institute. He observed that experts are experts by virtue of domain-specific problem-solving strategies together with a great deal of domain-specific knowledge. This view received support from research into the psychology of problem solving that suggested that expert problem-solving performance was attributable to the possession of domain-specific facts and rules (Chi et al., 1988).

The realisation that knowledge lay at the heart of expertise triggered a flurry of interest in knowledge elicitation and representation. Knowledge engineers soon discovered, however, that acquiring sufficient high-quality knowledge from individuals to build a robust and useful system was a very time-consuming and expensive activity. It seemed to take longer to elicit knowledge from experts than to write the expert system software. This problem became widely recognised as the *knowledge acquisition bottleneck* (Hayes-Roth et al., 1983), and it spawned an interest in the development, evaluation and practical application of a broad range of knowledge elicitation techniques that continued throughout the 1980s and 1990s.

Today, the scope of knowledge engineering efforts is much broader than simply the development of expert systems. With the advent of the web and Semantic Web,<sup>‡</sup> the focus of many knowledge engineering efforts has changed (Gil, 2011; Schreiber, 2013), and the development of formal computational ontologies<sup>§</sup> is now a major focus of attention for those concerned with the elicitation, representation and exploitation of human knowledge. There is also a broader recognition of the role that knowledge elicitation can play in corporate knowledge management. There are many different characterisations of knowledge management, but the central assumption is that knowledge is a valuable asset that must be managed (Nonaka and Takeuchi, 1995; Stewart, 1997). What we are looking for in knowledge management is a means to get the right knowledge to the right people at the right time and in the right form. These are difficult challenges, and many of them are identical to those encountered with the attempt to develop early knowledge-based systems (Hayes-Roth et al., 1983). There is thus a growing appreciation of the value of incorporating knowledge elicitation techniques into knowledge management initiatives, and it has been suggested that the tools, techniques, methods and approaches of knowledge engineering are well suited to the knowledge management enterprise (Gavrilova and Andreeva, 2012; Milton et al., 1999). One topic of particular interest concerns the use of knowledge elicitation techniques to support the transformation of tacit knowledge into explicit knowledge as part of the cycle of organisational knowledge creation (Nonaka and Takeuchi, 1995). Many of the knowledge elicitation techniques presented

\* Expert systems are computer programs that embody domain-specific knowledge and that perform at the same level as human experts within some domain (although they do not necessarily solve problems in the same way as human experts).

<sup>†</sup> Some early examples of such systems are MYCIN (Shortliffe, 1976) for diagnosing bacterial infections and PROSPECTOR (Duda et al., 1979) for supporting decisions relating to geological exploration.

<sup>‡</sup> The Semantic Web is a set of technologies that provide a common framework for the representation and exchange of knowledge and data in the context of the World Wide Web (Berners-Lee et al., 2001; Shadbolt et al., 2006).

<sup>§</sup> A 'computational ontology', in this case, is a formal, machine-readable representation of knowledge in some domain of interest. In the context of the Semantic Web, ontologies are typically created using the representational formalisms provided by the family of languages that goes under the heading of the Web Ontology Language or OWL. Such languages have both a formal semantics and an RDF/XML-based serialization. The formal semantics provide the basis for forms of machine-based reasoning in which a system is able to infer additional information based on the data that are explicitly represented, while the RDF/XML-based serialisation enables knowledge to be published and exploited within the distributed infrastructure of the World Wide Web.

later in this chapter can assist with this process, and they may thus play important roles in enabling organisations to realise their innovative potential.

This chapter will discuss the problem of knowledge elicitation for knowledge-intensive systems in general. These systems now come in a bewildering range of forms, from conventional expert systems through to intelligent tutoring systems, adaptive interfaces and workflow support tools. In many cases, the goal of knowledge elicitation is simply to generate representations of knowledge that may or may not be exploited in the context of computerised systems. One of the aims of knowledge elicitation, for example, may be to document the work-related knowledge and expertise that has developed within an organisation over a period of time. In addition, there may be a requirement to capture the knowledge of individuals who are about to leave an organisation or who have recently retired. These kinds of knowledge elicitation efforts often form part of an effort to preserve organisational knowledge and expertise by making the knowledge available to new recruits.

Another goal of knowledge elicitation and modelling, especially in more recent times, is to create computational ontologies that can be used in the context of the Semantic Web. The Semantic Web is a vision of how information can be represented and exchanged in the distributed computing environment of the World Wide Web. The essential idea is that information should be represented in a common form and with common semantics. This enables data to be shared, reused and processed across application, enterprise and community boundaries. Unlike the case with the conventional web, which is designed largely for human consumption, the aim of the Semantic Web is to support greater levels of machine intelligence and more advanced forms of human–machine interaction. In this respect, it is important to bear in mind that the Semantic Web is not a replacement for the conventional web; rather, it is something that sits alongside the conventional web and extends the range of capabilities and forms of interaction that can be delivered:

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. (Berners-Lee et al., 2001)

Ontologies play an important role in the context of the Semantic Web. They provide machine-readable representations of human knowledge that specify the knowledge structures of interest in some domain. Such forms of knowledge representation may serve a variety of purposes. As is the case with any form of web-accessible content, it is not always easy to anticipate the kind of ways in which these epistemic resources will be exploited. They may be used to support the implementation of intelligent systems, they may be used to support data interoperability and exchange solutions or they may simply be used to enable semantic search through domain-specific resource repositories.

Many problems arise before the elicitation of detailed domain knowledge can begin. First, we need to fully understand the goal of a knowledge engineering project. Sometimes a key failure is in formulating the role of a knowledge-based system; on other occasions, it is a failure to appreciate what it is realistic to build. Systems can fail because no one has thought of the social and organisational problems that must be resolved in deploying such a system. Very often, the effort and resources required to build systems are underestimated – this occurs in both the development and maintenance of systems. A particularly difficult situation arises when one is expected to conjure up knowledge for areas in which no evidence of systematic practice exists at all. Here, one is expected to provide theories for domains where there is no theory.

In terms of the actual process of knowledge elicitation, one may be able to gather information from a variety of non-human resources: textbooks, technical manuals, case studies and so on. However, in most cases, one needs to consult a practising expert. This may be because there is no documentation available, or because real expertise derives from practical experience in the domain rather than from a reading of standard texts. Few knowledge-intensive systems are ever built without recourse to experts at some stage. Those systems not informed by actual expert understanding



and practice are often the poorer for it. One of the recent slogans to emerge from the knowledge and cognitive engineering community is that the ‘gold is not in the documents’:

The gold is not in the documents. Document analysis is useful in bootstrapping researchers into the domain of study...but experts possess knowledge and strategies that do not appear in documents and task descriptions. Cognitive engineers invariably rely on interactions with experts to garner implicit, obscure, and otherwise undocumented expert knowledge. (Hoffman and Lintern, 2006, p. 215)

Given the need for expert involvement, it is typically the case that a knowledge engineer will be responsible for eliciting the expertise of experts. The main challenge here is to find a means by which the expert is enabled to communicate their knowledge to the person responsible for developing a knowledge solution. How can we establish the conditions that enable the expert to communicate the knowledge that underlies their expertise? This is a hard enough problem in itself, but there are a variety of circumstances that contrive to make the problem even harder. Much of the power of human expertise lies in laid-down experience, gathered over a number of years and represented as heuristics.\* Often, the expertise has become so routinised that experts no longer know how they accomplish particular tasks. In other cases, the knowledge required to build a system is distributed across an organisation and resides in the minds of a number of experts.

Of course, it is not just the capacity to elicit knowledge from an expert that is important. We would also like the knowledge elicitation process to be highly efficient and address the aforementioned knowledge acquisition bottleneck. Ideally, we would like to be able to use techniques that minimise the effort spent in gathering, transcribing and analysing an expert’s knowledge. We would also like to minimise the time spent with expensive and scarce experts. And, of course, we would like to maximise the yield of usable knowledge.

These sorts of issues lie behind the development of the many knowledge elicitation techniques that have become available over the past 20–30 years. A number of surveys of these techniques are now available (Cooke, 1994, 1999; Hoffman, 1987, 1989; Hoffman et al., 1995; Milton, 2012; Shadbolt, 2005; Shadbolt and Burton, 1995), and the current chapter builds on these existing surveys. We begin by describing, in sufficient detail for the reader to apply them, examples of major knowledge elicitation techniques. We then consider the features of domain experts and their associated expertise that are likely to directly affect the knowledge elicitation process. We also describe some of the issues that surround the appropriate selection of knowledge elicitation techniques as part of a programme of knowledge elicitation. Our attention then turns to some of the available software tools that support the knowledge elicitation process, typically by providing computerised versions of one or more knowledge elicitation techniques. Finally, we discuss some of the implications of the web and Semantic Web for knowledge elicitation efforts.

## KNOWLEDGE ELICITATION TECHNIQUES

There are a range of techniques that can be used to elicit knowledge from domain experts. The techniques we will describe are methods that we have found in our previous work to be both useful and complementary to one another. We can subdivide them into *natural* and *contrived* methods. The distinction is a simple one. A method is described as natural if it is one an expert might informally adopt when expressing or displaying expertise. Such techniques include interviews or the observation of actual problem solving. There are other methods we will describe in which the expert undertakes a contrived task. Examples here include concept sorting and the repertory grid technique. In the case of contrived tasks, the task elicits expertise in ways that are not usually familiar to an expert, and experts may feel uncomfortable when asked to perform them. Indeed, experts may feel they are performing badly with such methods, and they may question the value of such methods in

---

\* A heuristic is defined as a rule of thumb or generally proven method to obtain a result given particular information.

tapping into their expertise. In this respect, it is worth noting that we have found that an expert's own opinion of the worth of a technique is no guide as to its actual value (Schweikert et al., 1987). In addition, contrived techniques can sometimes prove more efficient than their non-contrived counterparts when it comes to knowledge elicitation (Burton et al., 1990). For these reasons, it is often useful to incorporate the use of contrived techniques into a programme of knowledge elicitation, although time will often be required to explain the use of these techniques to domain experts.

## INTERVIEWS

Almost everyone starts in knowledge elicitation by determining to use an interview. The interview is the most commonly used knowledge elicitation technique, and it takes many forms. Three kinds of interview are generally recognised within the knowledge engineering community. These are the *unstructured*, *semi-structured* and the *structured* interview. In all cases, the main aim of the interview is to elicit information regarding how a particular task is performed or how a particular decision is made.

The starting point for most new knowledge engineering efforts will be an unstructured interview since this is the best means of establishing rapport between the knowledge elicitor and the expert. In addition, unstructured interviews provide a useful means of 'bootstrapping' the elicitor's understanding of the target domain – they provide an opportunity for the elicitor and the expert to discuss the domain in an informal setting with no constraints as to what can be discussed. Unfortunately, this is also one of the main drawbacks of the unstructured interview. By virtue of being unstructured, the interview can easily allow the elicitor and expert to dwell on irrelevant topic areas or cover important areas in insufficient depth. For these reasons, there is often a requirement to resort to more structured interviewing methods.

The structured interview is a formal version of the interview in which the person eliciting the knowledge plans and directs the session.\* A significant benefit of the structured interview is that it provides structured transcripts that are easier to analyse than unstructured conversations. This serves to improve the efficiency of the structured interview, and it also enables the elicitor and expert to focus their attention on a limited subset of important topics.

Although it is common to see the structured interview as a single technique, it is probably best to think of it as a class of techniques (Hoffman et al., 1995). There are, in fact, many varieties of structured interviews. In *forward scenario simulation interviews*, for example, the expert is walked through the problem verbally by the elicitor who presents decision- or task-relevant information to the expert, and the expert is asked to respond accordingly (Cordingley, 1989; Grover, 1983). Another kind of structured interview is the *fixed probe interview* in which specific probe questions are used to elicit domain knowledge. A template for such an interview is as follows:

1. Ask the expert to give a brief (10 min) outline of the target task, including the following information:
  - a. A description of the possible solutions or outcomes of the task
  - b. A description of the variables that affect the choice of solutions or outcomes
  - c. A list of the major rules or procedures that connect the variables elicited to the solutions or outcomes
2. Take each rule or procedure elicited in Stage 1, ask when it is appropriate and when it is not, and if it is a procedure, ask how it is performed. The aim is to reveal the scope (generality and specificity) of each existing rule and hopefully generate some new rules.
3. Repeat Stage 2 until it is clear that the expert will not produce any additional information.

---

\* In practice, we have found that it is often useful to involve the expert in the planning of a structured interview session. Expert input at the planning stage can be useful in terms of identifying important areas, and it also enables the expert to have an understanding of what topics will be discussed in advance of the knowledge elicitation session.

A useful way of obtaining a domain overview (Stage 1 of the structured interview) is to ask probe questions that relate to an individual's specific experience. It is also important in this technique to be specific about how to perform Stage 2. We have found that it is helpful to constrain the elicitor's interventions to a specific set of *probes*, each with a specific function. Here is a list of probes (P) and functions (F) that can help in the first two stages of the interview.

*P1.1:* Could you tell me about a typical case?

*F1.1:* Provides an overview of the domain tasks and concepts.

*P1.2:* Can you tell me about the last case you encountered?

*F1.2:* Provides an instance-based overview of the domain tasks and concepts.

*P2.1:* Why would you do that?

*F2.1:* Converts an assertion into a rule.

*P2.2:* How would you do that?

*F2.2:* Generates *lower-order* rules.

*P2.3:* When would you do that? Is <the rule> always the case?

*F2.3:* Reveals the generality of the rule and may generate other rules.

*P2.4:* What alternatives to <the prescribed action/decision> are there?

*F2.4:* Generates more rules.

*P2.5:* What if it were not the case that <currently true condition>?

*F2.5:* Generates rules for when current condition does not apply.

*P2.6:* Can you tell me more about <any subject already mentioned>?

*F2.6:* Used to generate further dialogue if the expert dries up.

*P2.7:* Can you tell me about an unusual case you encountered/heard about from some other expert?

*F2.7:* Refines the knowledge to include rare cases and special procedures.

The idea here is that the elicitor engages in a type of slot/filler dialogue. The provision of template questions about concepts, relations, attributes and values makes the elicitor's job much easier. It also provides sharply focused transcripts that facilitate the process of extracting usable knowledge. Of course, there will be instances when none of these probes are appropriate (such as the case when the elicitor wants the expert to clarify something). However, you should try to keep these interjections to a minimum. The point of specifying such a fixed set of linguistic probes is to constrain the expert to giving you all, and only, the information you want.

The following sample of dialogue is taken from a real interview of this kind. It is the transcript of an interview by a knowledge engineer (KE) with an expert (EX) in the domain of geological analysis.\*

*KE:* What would you do at this stage?

*EX:* I would look at the grain size of the hand specimen and see how fine it was.

*KE:* Why would you look at the grain size?

*EX:* That will tell me if the rock has been formed near to the surface or deep inside the earth. The finer the grain size the faster it cooled. Coarse crystals indicate that the rock was cooling slowly + forming deeper down + we say its emplacement is plutonic + if it cooled near the surface its emplacement is volcanic.

*KE:* Are there any alternatives to coarse and fine grain size?

*EX:* There are glasses + you can't see any structure here because the rock cooled so fast.

*KE:* What would you look at next?

*EX:* Colour is important + the lighter the rock the more acidic it is.

*KE:* Why is a lighter rock more acidic?

*EX:* Acidic rocks are higher in quartz and colour is a good indicator of quartz content – leucocratic or light things have a lot of quartz – melanocratic that is darker rocks have olivines and pyroxenes.

---

\* In the transcripts, we use the symbol + to represent a pause in the dialogue.

This is quite a rich piece of dialogue. From this section of the interview alone, we can extract numerous rules such as:

```
IF grain size is large
THEN rock is plutonic

IF rock is leucocratic
THEN rock has high quartz content
```

Of course, these rules may need refining in later elicitation sessions, but the text of the dialogue shows how the use of the specific probes has revealed a well-structured response from the expert.\*

Techniques exist to impose a lesser amount of structure on an interview. These kinds of techniques can be referred to as types of semi-structured interview. One example of a semi-structured interview is the *knowledge acquisition grid* (LaFrance, 1987). This is a matrix of knowledge types and forms – examples of knowledge forms are *layouts* and *stories*, while some examples of question types are *grand tour* and *cross-checking*. A grand tour involves such things as distinguishing domain boundaries and the overall organisation of goals; cross-checking involves the engineer attempting to validate the acquired knowledge by, for example, playing devil's advocate.

Another form of semi-structured interview technique is the *teachback technique* of Johnson and Johnson (1987). In this technique, the expert explains something to the elicitor who then attempts to explain it to the expert – the knowledge is effectively 'taught back' to the expert. The expert then has an opportunity to check and, if necessary, amend the information.

Unstructured interviews have no agenda (or, at least, no *detailed* agenda) set either by the knowledge elicitor or by the expert. Of course, this does not mean that the elicitor has no goals for the interview, but it does mean that she has considerable scope for proceeding. As mentioned earlier, the unstructured interview is useful for a variety of reasons. First, the approach can be used whenever one of the goals of the interview is to establish a rapport between the expert and the knowledge elicitor. There are no formal barriers to the discussion covering whatever material either participant sees fit. Second, one can get a broad view of the topic easily; the knowledge elicitor can 'fill in the gaps' in her own perceived knowledge of the domain. Third, the expert can describe the domain in a way with which he is familiar, discussing topics that he considers important and ignoring those he considers uninteresting.

The disadvantages are clear enough: the lack of structure can lead to inefficiency; the expert may be unnecessarily verbose; the expert may concentrate on topics whose importance he exaggerates; the coverage of the domain may be patchy; and the data acquired may be difficult to integrate, either because it does not form a coherent body of content or because there are inconsistencies (this will be even more likely if the information provided by *several* experts is to be collated).

In all of the interview techniques mentioned so far (and in some of the other techniques as well), there exist a number of dangers that have become familiar to practitioners of knowledge elicitation. One problem is that in an interview, experts will only produce what they can verbalise. If there are non-verbalisable aspects to the domain, the interview will not recover them. It may be that the knowledge was never explicitly represented or articulated in terms of language (consider, for example, pattern recognition expertise). Then, there is the situation where the knowledge was originally learnt explicitly in a propositional or language-like form. However, in the course of experience, such knowledge has become routinised or automatised.† This can happen to such an extent that experts may regard the complex decisions they make as based only on hunches or intuitions. In actual fact, these decisions are based upon large amounts of remembered data and experience and

\* In fact, a possible second-phase elicitation technique would be to present these rules back to the expert and ask about their validity, scope and so forth.

† We often use a computing analogy to refer to this situation and speak of the expert as having *compiled* the knowledge.

the continual application of that knowledge. In this situation, they tend to give *black box* replies such as ‘I don’t know how I do that...’ or ‘It is obviously the right thing to do...’.

Another problem arises from the observation that people (and experts in particular) often seek to justify their decisions in any way they can. It is a common experience of the knowledge elicitor to get a perfectly valid decision from an expert and then to be given a spurious justification as to why it was made and how it originated.

For these and other reasons, one should always supplement interviews with additional elicitation methods. In general, knowledge elicitation should always consist of a programme of techniques and methods (see section ‘Methodologies and Programmes’).

## PROTOCOL ANALYSIS

Protocol analysis (PA) is a generic term for a number of different ways of performing some form of analysis of the expert(s) actually solving problems in the domain. In all cases, the elicitor takes a record of what the expert does using written notes or (preferably) an audio or video recording. Transcripts or protocols are then made from these records, and the elicitor tries to extract meaningful structure, rules and processes from the protocols.

We can distinguish two general types of PA: *online* and *offline*. In online PA, the expert is recorded solving a problem and *concurrently* a commentary is made. The nature of this commentary specifies two sub-types of the online PA method. The expert performing the task may be describing what they are doing as problem solving proceeds. This is called *self-report*. A variant on this is to have another expert provide a running commentary on what the expert performing the task is doing. This is called *shadowing*.

Offline PA allows the expert to comment retrospectively on the problem-solving session, usually by being shown an audio-visual record of it. This may take the form of a *retrospective self-report* by the expert who actually solved the problem. Alternatively, it may take the form of a retrospective report by another expert – this has recently been referred to as *collegial verbalisation* (Erlandsson and Jansson, 2007) – or there could be group discussion of the protocol by a number of experts including its originator. In situations where only a behavioural protocol (such as a video recording) is obtained, then some form of retrospective verbalisation of the problem-solving episode will obviously be required.

In many cases, the focus of PA is on verbal data. In this case, the technique is typically referred to as verbal PA (see Bainbridge and Sanderson, 2005). Other types of events, such as eye movements, gestures and other non-verbal behaviours may also be the focus of PA, although this is rarely seen in practice. Combining the analysis of, for example, eye movements with verbal reports may be useful in some cases, particularly in situations where the aim is to better understand the allocation of attention to particular environmental cues and sources of task-relevant information. In one study, for example, Van Gog et al. (2005) used a combination of eye movement data and concurrent verbal PA in order to explore expertise-related differences in electrical circuit troubleshooting performance.

In deciding between the various kinds of PA technique on offer, it is worth bearing in mind a number of issues. First, in their classic treatment of PA, Ericsson and Simon (1996) recommend the use of concurrent verbal reports (i.e. online self-reports) over retrospective ones. One of the possible problems with retrospective reports is that the conditions associated with verbalisation in the two cases may differ, and this may affect information processing accordingly. In general, it is assumed that the longer the delay between performance and report, the greater this problem becomes. As a result, it is predicted that more immediate retrospective reports are the most similar to concurrent ones. On the other hand, concurrent verbalisation techniques can present a number of problems for experts, such as interference with the execution of skilled actions. Ericsson and Simon (1996) suggest a number of conditions under which verbal report procedures should succeed or fail.

For instance, verbal reports are not as effective for eliciting knowledge when the problem is novel or the reporter has low verbal ability or is inhibited in some way. When these sorts of conditions are encountered in the context of a programme of knowledge elicitation, it may be beneficial to incorporate more retrospective PA techniques.

In addition to decisions about the choice between online and offline PA, decisions also have to be made about the extent to which other experts (other than those actually performing the task) are involved in the verbal commentary. Typically, the individual performing the task provides the verbal report, either concurrently or retrospectively. However, other techniques, such as that of collegial verbalisation,\* have also been the focus of recent attention (Erlandsson and Jansson, 2007, 2013). One issue of interest here concerns the extent to which the reports provided by other experts match those provided by the performing expert. In one study comparing collegial verbalisation with retrospective self-report, Erlandsson and Jansson (2013) found a number of similarities between the protocol data delivered by the two techniques, suggesting that collegial verbalisation may be as effective as retrospective self-report. Clearly, in a situation where a video record of expert performance is available, a number of protocols can be obtained using multiple experts. This may serve to improve the reliability and completeness of the resulting knowledge base.

In trying to decide when it is appropriate to use PA, bear in mind that it is alleged that different knowledge elicitation techniques differentially support the elicitation of particular kinds of information. This is commonly known as the *differential access hypothesis* (Hoffman et al., 1995). With PA, it is claimed that the sorts of knowledge elicited include the ‘when’ and ‘how’ of using specific knowledge. It can reveal the problem-solving and reasoning strategies, evaluation procedures and evaluation criteria used by the expert, and procedural knowledge about how tasks and sub-tasks are decomposed. A PA gives you a complete episode of problem solving. It can be useful as a verification method to check that what people say is actually what they do. It can also take you deeper into a particular problem. It is, however, intrinsically a narrow method since it can be used to analyse only a relatively small number of problems within the domain.

Before PA sessions can be held, a number of preconditions should be satisfied. The first of these is that the elicitor is sufficiently acquainted with the domain to understand the expert’s tasks. Without this, the elicitor may completely fail to record or take note of important parts of the expert’s behaviour.

A second requirement is the careful selection of problems for PA. This sampling of problems is crucial. PA sessions may take a relatively long time, and usually only a few problems can be addressed in any programme of acquisition (Shadbolt and Burton, 1989). Therefore, the selection of problems should be guided by how representative they are. Asking experts to sort problems into some form of order (Chi et al., 1981, 1982) may give an insight into the classification of types of problems and help in the selection of suitable problems for PA (see also the following sections on concept sorting, repertory grids and ladder grids for methods that can be used to help classify and structure problems).

A further condition for effective PA is that the expert(s) should not feel embarrassed about describing their expertise in detail. It is preferable for them to have experience in thinking aloud. Uninhibited thinking aloud has to be learned in the same way as talking to an audience. One or two short training sessions may be useful. In these training sessions, a simple task, such as long multiplication, can be used as an example. This puts the expert at ease and familiarises them with the task of talking about their problem solving.

In order to collect protocols, the expert is asked to ‘think aloud’ while performing some task, and the resulting commentary is typically recorded and transcribed. In terms of recording techniques,

---

\* Collegial verbalisation is based on the procedure of videotaping practitioners while they perform their normal work tasks in their normal work setting. This is followed up by having a close colleague of the practitioner watch the video recordings and verbalise.



it is preferable to use video recordings rather than audio recordings. This is because video recordings capture more information about the context in which problem solving occurs, which can help to support the resulting analysis. In particular, the following two advantages of video recording techniques have been noted by Bainbridge and Sanderson (2005):

1. First, video recordings often help to disambiguate what is being referred to in the case of situated forms of problem-solving activity. Subjects often make use of pronouns, such as ‘when *it’s* at 55’, and the presence of a visual record can help to disambiguate what is being referred to. Also, as noted by Bainbridge and Sanderson (2005), video recordings can help when people use general anaphoric references supplemented by pointing; for example ‘*that* is too high so I’ll lower *this* until it is between *these*’.
2. A second advantage of video recording techniques relates to the fact that is often useful to have information about the total task environment in which problem solving occurs. This can be used at a later time to assess to what extent people’s behaviour is influenced by features of the environment that are not explicitly mentioned in the verbal report.

One of the main drawbacks of video recording techniques is, of course, the amount of data they make available for analysis. It can be difficult to avoid the temptation to scale up the analytic effort when confronted with such detailed records, and discipline is often required to limit attention to information of relevance to the knowledge elicitation effort.

When actually conducting a PA, the following are a useful set of tips to help enhance its effectiveness:

1. Present the problems and data in a realistic way. The way problems and data are presented should be as close as possible to a real situation.
2. Transcribe the protocols as soon as possible. The meaning of many expressions is soon lost, particularly if the protocols are not recorded.
3. Avoid long self-report sessions. Because of the need to perform a double task – combining expert performance with verbal commentary – the process of thinking aloud is significantly more tiring for the expert than being interviewed. This is one reason why shadowing is sometimes preferred.
4. In general, the presence of the elicitor is required in a PA session. Although the elicitor adopts a background role, her very presence suggests a listener to the interviewee and lends meaning to the talking aloud process. Therefore, comments on audibility, or even silence by the elicitor, are quite acceptable.

When a verbal or behavioural transcript has been obtained, we next have to undertake its analysis. A number of approaches to the analysis of verbal protocols have been described in previous work, such as that by Bainbridge and Sanderson (2005). In general, however, it is acknowledged that there are no objective-independent techniques for doing these analyses, and this means that analysts ‘have to use both their own natural language understanding processes, and their knowledge of the task, in order to make sense of what is going on, to infer missing passages, and to interpret the results of summary analyses’ (Bainbridge and Sanderson, 2005, p. 166). For the purposes of most knowledge elicitation exercises, the analysis will typically involve the ‘encoding’ of the protocol transcript into ‘chunks’ of knowledge (actions, assertions, propositions, keywords, etc.), and it should result in a rich domain representation with many elicited domain features together with a number of specified links between those features. The following example is from a self-report of an expert geologist. It is immediately apparent that protocols can be extremely dense sources of information. A very significant amount of work



is required to analyse and structure the content in this very small fragment of a self-report concerning one rock specimen.

To start off with it's obviously a fairly coarse-grained rock... and you've got some nice big orthoclase crystals in here – this is actually SHAP GRANITE – I know it just because everybody's seen SHAP GRANITE – or it's a very strong possibility that it's SHAP GRANITE... it's a typical teaching specimen – as I say the obvious things are these very big orthoclase crystals pink colouration and you can certainly see some cleavage in some of them – you can certainly make out there are feldspar cleavages in there – it's a coarse-grained rock anyway, you can see the crystals nice and coarsely – these large porphyritic crystals – you can see, in the ground mass, you can see quartz – get some light on it (HOLDS SPECIMEN UP TO WINDOW) quartz, which is this fairly clear mineral you can actually look into it and see through it as opposed to calcite or feldspars where it's more cloudy – you can't actually see any good crystal faces on these cut sections – small flakes of biotite, black micaceous looking – small plates, you can certainly see some on this specimen even without a hand lens.

There are a number of principles that can guide the PA. For example, analysis of the verbalisation resulting in the protocol can distinguish between information that is attended to during problem solving, and that which is used implicitly. A distinction can be made between information brought out of memory (such as a recollection of a similar problem solved in the past) and information that is produced 'on the spot' by inference. The knowledge chunks referred to earlier can be analysed by examining the expert's syntax, or the pauses he takes, or other linguistic cues. Syntactical categories (e.g. use of nouns, verbs) can help distinguish between domain features and problem-solving actions, etc., in general, for multiple analysts to perform the encoding independently. This provides insight into the reliability of certain forms of encoding, and it also serves to highlight areas of contention that may need to be the focus of future knowledge elicitation sessions.

The focus and depth of the analytic efforts are typically dictated by the goals of the knowledge elicitation exercise. If the aim is to understand the sequential ordering of tasks in the context of some larger business process, this will require a less detailed form of PA compared to situations where the aim is to develop a computational model of the mental processes associated with problem-solving behaviour.

When appropriately elicited, verbal and non-verbal protocols can help to illuminate the normal sequential flow of working and thinking, and they are thus valuable components of the analyst's knowledge elicitation toolkit. In spite of this, PA does have its limitations. First, PA techniques share with the unstructured interview the problem that they may deliver unstructured transcripts that are hard to analyse. Moreover, they focus on particular problem cases, and so the scope of the knowledge produced may be very restricted. It is difficult to derive general domain principles from a limited number of protocols. These are some of the practical disadvantages of PA. However, there are more subtle problems. For example, two actions, which look exactly the same to the knowledge elicitor, may be very different in their extent and intent. For example, our geologist who applies a particular test to a specimen may apply that same test to another but with a quite different purpose. The knowledge elicitor simply does not know enough to discriminate the actions.

Another source of concern stems from the possibility of distorted information – the risk that PA may yield information that is not an accurate reflection of what takes place in task settings where the technique is not being employed. The causes of these distortions are outlined by Bainbridge and Sanderson (2005). They include:

1. The fact that being asked to give a verbal protocol changes the nature of the task that is being performed. A task that typically involves a number of concurrent actions may instead be performed in a sequential fashion as a result of the constraints imposed by the need to verbalise what one is doing. In cases where there are multiple ways of accomplishing a task, an expert may resort to a method that is easier to verbalise. Self-report techniques may also interfere

with expert performance. There is some empirical evidence that attending to the components of a well-learned skill can impair performance (Beilock et al., 2002; Gray, 2004), and it thus seems likely that by asking an expert to think aloud we are changing the nature of the task being performed. Some cases of skilled performance are probably best demonstrated when the expert is left to perform the task automatically without the kind of attentional reorganisation that is required by PA. This may also be the case with certain types of decision-making expertise. By asking the expert to verbalise, one is, in some sense, destroying the point of doing PA – to access procedural, real-world knowledge.

2. The temporal constraints involved in giving a verbal protocol. In situations where people are working under time constraints, there may be limits to what people can verbalise. In particular, there may be insufficient time to report task-relevant information that is brought to mind and then quickly forgotten as a result of the tempo of task performance.
3. The fact that giving a self-report is a socially situated activity involving self-presentation issues. People may, for example, want to appear to rational and knowledgeable to a professional observer, and this may influence the content of the self-report accordingly.
4. The fact that some aspects of the task may be performed automatically and the expert may not have conscious access to the knowledge that is being used. This is particularly the case with tasks involving advanced perceptuo-motor skills.
5. The limited scope of the technique. By focusing on a limited number of tasks, PA may inadequately sample the total knowledge possessed by an expert. As noted by Bainbridge and Sanderson (2005), 'knowledge about the components, mechanisms, functions and causal relations in a machine, memories of specific events, and helpful categories will be mentioned explicitly only if the task involves some problem solving that requires the person to review this sort of evidence' (p. 162).

Having pointed to these drawbacks, it is also worth remembering that context is often important for memory – and hence for problem solving. For most non-verbalisable knowledge, and even for some verbalisable knowledge, it may be essential to observe the expert performing the task in a naturalistic setting. It may be that this is the only situation in which the expert is actually able to demonstrate their expertise.

## CRITICAL DECISION METHOD

The critical decision method (CDM) is 'a retrospective interview strategy that applies a set of cognitive probes to actual non-routine incidents that required expert judgement or decision making' (Klein et al., 1989, p. 464). As a knowledge elicitation technique, the CDM contains elements of both interviewing and PA but in a context that stresses the examination of problem solving in naturalistic decision-making contexts (Zsombok and Klein, 1997). The technique involves the expert being guided through the recall and elaboration of previously encountered cases, especially ones that were, in some sense, unusual, difficult or otherwise involved critical decisions. Such cases are often particularly memorable for the domain expert, and this serves as an aid to the elicitation of important information, such as the information the expert needs to make decisions in particular contexts. At the same time, incidents that are difficult or non-routine are typically ones that provide the richest source of information about the knowledge and capabilities of domain experts. Detailed presentations of this method, along with summaries of studies illustrating its use, can be found in Klein et al (1989), Crandall et al. (2006), O'Hare et al. (1998) and Hoffman et al. (1998).

As originally presented by Klein et al. (1989), a CDM session is organised into five steps:

1. *Select incident.* In the first step, the expert is guided in the recall and recounting of a specific incident and its associated context. As mentioned earlier, the aim is to select an incident that is unusual or non-routine. The expert may be asked to 'select an incident that

was challenging and that, in his or her own decision making, might have differed from someone with less experience' (Klein et al., 1989, p. 466). As a second example, experts may be asked to focus on incidents that are 'in some manner unusual and difficult (i.e. where the best choice was not clear cut) in which the [expert] felt that their expertise and experience made a critical difference to the outcome' (O'Hare et al., 1998, p. 1700).

2. *Obtain unstructured incident account.* In the second step, the expert is asked to describe the incident from their own perspective. This step accomplishes a number of goals. First, it provides the basis for an analyst's initial understanding of the incident in question. Second, it serves to activate the expert's memory of an incident as the basis for subsequent questioning.
3. *Construct incident timeline.* After the incident has been described by the expert, a timeline of the account is constructed. This serves to establish the sequence and duration of each event reported by the expert.
4. *Decision point identification.* Once a timeline has been constructed, decision points in the timeline are identified and specific decisions are marked for further probing. In general, decisions are subjected to further probing if the expert feels that additional courses of action are possible, or if another expert might have chosen a different course of action.
5. *Decision point probing.* Any decision points that were marked for further probing in step 4 are analysed in more detail using a set of cognitive probes.

Table 7.1 contains a range of probe question types with exemplars that we have found to be particularly useful when applying the CDM. Although these are typically used in step 5 of the CDM method, there is no reason why these questions cannot be used in the context of other steps. In addition, the probes listed in Table 7.1 do not exhaust the range of probes that could be used in the context of the CDM. O'Hare et al. (1998), for example, present an extended set of cognitive probes that are designed to 'obtain additional information on the perceptual and cognitive structures and processes that appear to mediate expertise' (p. 1700).

**TABLE 7.1**  
**Sample CDM Probe Questions**

| Probe Type           | Probe Examples                                                                                                   |
|----------------------|------------------------------------------------------------------------------------------------------------------|
| Cues                 | What were you seeing, hearing, smelling?                                                                         |
| Knowledge            | What information did you use in making this decision? How was it obtained?                                       |
| Analogues            | Were you reminded of any previous incidents?                                                                     |
| Scenarios            | Does this case fit a standard or typical scenario? Does it fit a scenario you were trained to deal with?         |
| Goals                | What were your specific goals and objectives at the time?                                                        |
| Options              | What other courses of action were considered or available?                                                       |
| Choice               | How was this option selected/other options rejected? What rule was being followed?                               |
| Anticipation         | Did you imagine the possible consequences of this action? Did you imagine the events that would unfold?          |
| Experience           | What specific training or experience was necessary or helpful in this decision? What more would have helped?     |
| Decision making      | How much time pressure was involved in making the decision? How long did it take to make the decision?           |
| Aiding               | What training, knowledge or information could have helped?                                                       |
| Situation assessment | If you were asked to describe the situation to a colleague at this point, how would you summarise the situation? |
| Errors               | What mistakes are likely at this point? How might a novice have behaved differently?                             |
| Hypotheticals        | If a key feature of the situation had been different, what differences would it have made in your decision?      |

The outcome of the CDM is a range of products which can be used to support training and system development activities (Klein et al., 1989). One of the most important products is referred to as the critical cue inventory (CCI), which is a collection of all the perceptual cues that are used to guide the consideration and selection of particular decisions. In the case of medical decision making, for example, the CCI could include a list of cues for recognising critical conditions, such as early signs of cardiopulmonary distress (see Klein et al., 1989). Another important product of the CDM is the situation assessment record (SAR). The SAR records the changes in goals and cue usage associated with situation assessment processes. It typically combines information about the cues being sought or identified, the expectancies generated by these cues, the goals activated by the current situation and the selected course of action resulting from knowledge about the assessed situation.

A typical CDM session can last around 2 h. Depending on the domain, much of this time may be spent recollecting a rich complex incident. In other settings, the majority of the effort may be devoted to examining counterfactual situations. The CDM does have its limitations. In distributed problem-solving situations, no one individual may handle more than one element of a task. The individuals, in this case, would never know whether their judgements or assessments were correct within the context of the larger socially distributed process. In addition, in high-workload environments, we have sometimes observed that incidents and events can become merged. When responding to an opening query, one sometimes sees an expert recount an incident but then become confused when asked for a timeline or other details. Despite these shortcomings, the style of interview and the attention paid to particular incidents often provide a rich output from which the elicitor can extract important task-relevant knowledge. An added bonus is that the case studies resulting from the application of the CDM can often serve as important training materials.

## CONCEPT SORTING

Unlike interview techniques and PA, concept sorting is a form of contrived knowledge elicitation technique that is likely to be unfamiliar to the domain expert. The technique is useful when we wish to elicit the different relationships that exist between a fixed set of concepts. In the version of concept sorting we describe here, an expert is presented with a number of cards on each of which is printed a concept word. The cards are shuffled, and the expert is asked to sort the cards into either a fixed number of piles or else to sort them into any number of piles the expert finds appropriate. This process is repeated many times.

Using this task, one attempts to get multiple views of the structural organisation of knowledge by asking the expert to do the same task over and over again. Each time the expert sorts the cards, he should create at least one pile that differs in some way from previous sorts. The expert should also provide a name or category label for each pile on each different sort. This is often referred to as the *dimension* along which concepts are sorted (see Table 7.3), and it typically identifies a particular property or attribute associated with a class of objects (e.g. 'grain size' may be represented as an attribute of the 'rock' class).

Performing a card sort requires the elicitor to have some basic conception of the domain. Cards have to be made with the appropriate labels before the session. However, no great familiarity is required as the expert provides all the substantial knowledge in the process of the sort. We now provide an example from our geology domain to show the detailed mechanics of a sort.

The concepts printed on a set of cards are the names of igneous rocks drawn from a structured interview with the expert. He had previously described 18 rock types, which are presented in Table 7.2.

The expert was shown possible ways of sorting cards in a *toy* domain as part of the briefing session. He was then asked to sort the real elements in the same way. The dimensions/piles which the expert used for the individual card sorts are presented in Table 7.3.

Table 7.4 shows the piles into which each of the rock types in Table 7.2 was placed as part of the sequence of card sorts. As can be seen from Table 7.4, many of the elements are distinguishable from one another, even with this limited number of card sorts.

**TABLE 7.2**  
**Names of 18 Types of Igneous Rock Elicited from a Geologist as Part of a Structured Interview**

|                 |                    |
|-----------------|--------------------|
| 1. Adamellite   | 10. Granite        |
| 2. Andesite     | 11. Lherzolite     |
| 3. Basalt       | 12. Microgranite   |
| 4. Dacite       | 13. Peridotite     |
| 5. Diorite      | 14. Picrite basalt |
| 6. Dolerite     | 15. Rhyodacite     |
| 7. Dunite       | 16. Rhyolite       |
| 8. Gabbro       | 17. Syenite        |
| 9. Granodiorite | 18. Trachyte       |

**TABLE 7.3**  
**Results of Seven Card Sorts Undertaken as Part of a Concept-Sorting Knowledge Elicitation Session with a Geologist**

| Sort # | Dimension            | Piles                                                            |
|--------|----------------------|------------------------------------------------------------------|
| 1      | Grain size           | 1 = Coarse, 2 = medium, 3 = fine                                 |
| 2      | Colour               | 1 = Melanocratic, 2 = mesocratic, 3 = leucocratic                |
| 3      | Emplacement          | 1 = Intrusive, 2 = extrusive                                     |
| 4      | Presence of olivine  | 1 = Always, 2 = possibly, 3 = never                              |
| 5      | Presence of quartz   | 1 = Always, 2 = possibly, 3 = never                              |
| 6      | Percentage of silica | 1 ≥ 68%, 2 ≤ 68%, 3 = about 68%                                  |
| 7      | Density              | 1 = Very light, 2 = light, 3 = medium, 4 = dense, 5 = very dense |

**TABLE 7.4**  
**Positioning of Cards Representing Different Types of Igneous Rock (See Table 7.2) in the Piles Resulting from Seven Card Sorts with a Geologist (See Table 7.3)**

| Sort | Rock |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |
|------|------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|
|      | 1    | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 1    | 1    | 3 | 3 | 3 | 1 | 2 | 1 | 1 | 1 | 1  | 1  | 2  | 1  | 3  | 3  | 1  | 1  | 3  |
| 2    | 3    | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 3 | 3  | 1  | 3  | 1  | 1  | 3  | 3  | 3  | 3  |
| 3    | 1    | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  | 2  | 2  | 2  | 1  | 2  |
| 4    | 1    | 3 | 2 | 3 | 3 | 2 | 1 | 2 | 3 | 3  | 1  | 3  | 1  | 1  | 3  | 3  | 3  | 3  |
| 5    | 1    | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 1 | 1  | 3  | 1  | 3  | 3  | 1  | 1  | 1  | 2  |
| 6    | 2    | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 3 | 1  | 2  | 1  | 2  | 2  | 1  | 1  | 2  | 2  |
| 7    | 1    | 3 | 4 | 2 | 3 | 4 | 5 | 4 | 1 | 1  | 5  | 1  | 5  | 4  | 2  | 1  | 3  | 2  |

Using the results of the card sorts, we can attempt to extract decision rules directly. An example of a rule extracted from the card-sorting data is as follows:

|      |                                  |                 |
|------|----------------------------------|-----------------|
| IF   | the grain size is fine           | (sort 1/pile 3) |
| AND  | the colour is mesocratic         | (sort 2/pile 2) |
| AND  | its emplacement is extrusive     | (sort 3/pile 2) |
| AND  | it does NOT contain olivine      | (sort 4/pile 3) |
| AND  | may possibly contain quartz      | (sort 5/pile 2) |
| AND  | it contains less than 68% silica | (sort 6/pile 2) |
| AND  | its density is medium            | (sort 7/pile 3) |
| THEN | the rock is andesite             | (outcome 2)     |

As can be seen from this example, card sorts often produce long and cumbersome rules. In fact, many of the clauses may be redundant. For example, once you have established that the grain size is small, then it is going to be an extrusive rock. The utility of the technique, however, does not reside solely in the production of decision rules. We can use it, as we have said, to explore the general inter-relationships between concepts in the domain. We can also use the technique to elicit the features of concepts\* that might not otherwise surface in the context of other techniques.

The advantages of concept sorting can be characterised as follows. It is fast to apply and easy to analyse. It also serves to make explicit the implicit structure that experts impose on their expertise. In fact, the process of performing concept sorting is often instructive to the expert – a sort can lead the expert to see structure that he himself has not consciously articulated before. Concept sorting can also be a highly efficient technique, especially when computerised support is available for the implementation and analysis of the sorting procedure. Unlike the case with interviews and PA, time can often be saved by not having to transcribe and analyse lengthy verbal reports.† Finally, in domains where the concepts are perceptual in nature (i.e. x-rays, layouts and pictures of various kinds), the cards can be used as a means of presenting these images and attempting to elicit names for the categories and relationships that might link them.

The technique does, of course, have its disadvantages. Experts can often confound dimensions by not consistently applying the same semantic distinctions throughout an elicitation session. Alternatively, they may oversimplify the categorisation of elements, missing out important caveats.

## REPERTORY GRIDS

This technique has its roots in the psychology of personality (Fransella et al., 2003; Jankowicz, 2003; Kelly, 1955). It is designed to reveal a conceptual map of a domain in a fashion similar to the concept-sorting technique discussed earlier. The work of Mildred Shaw and Brian Gaines was particularly important in promoting the use of the technique (Shaw and Gaines, 1987), and the development of computerised versions of the technique was an important step in making the repertory grid a standard element of the knowledge elicitation technique palette (the technique as developed in the 1950s was very time consuming to administer and analyse by hand). One example of repertory grid software is WebGrid 5, which can be accessed from the WebGrid website.‡ WebGrid 5 is the latest

\* It is important to bear in mind that although the name of the technique suggests that its use is limited to concepts, the technique can, in fact, be applied to knowledge elements of any type. The cards used in a card-sorting task, e.g., might name tasks, goals, actions resources and so on; the only restriction is that in any sorting session, the cards should be of the same knowledge type.

† Although it is not necessary to make an audio recording of concept sorting sessions, we recommend that such records are, in fact, made. An expert makes many asides, comments and qualifications in the course of sorting ranking and so on. In fact, one may choose to use the contrived methods as a means to carry out auxiliary structured interviews. The structure this time is centred on the activity of the technique.

‡ See <http://gigi.cpsc.ucalgary.ca/>.

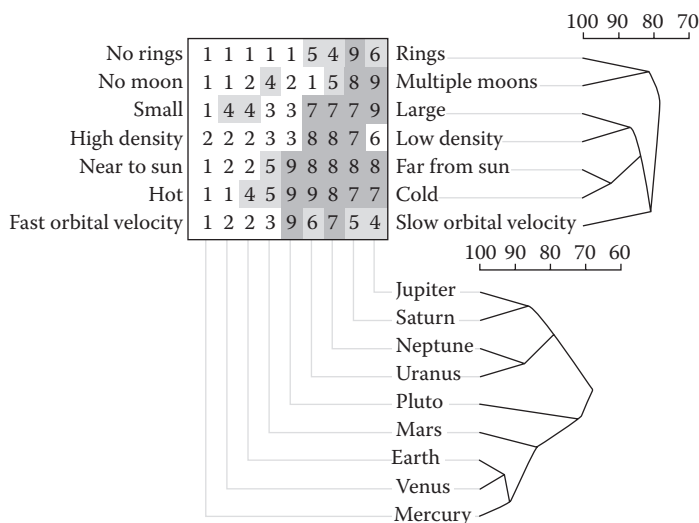


version of the web-based implementation of the repertory grid technique that was described by Gaines and Shaw (Gaines and Shaw, 1997; Shaw and Gaines, 2001), as part of their attempt to make knowledge acquisition technologies accessible via the World Wide Web. The software provides an excellent means of experimenting with the approach and indeed undertaking machine-supported elicitation sessions.

As part of the repertory grid technique, subjects are presented with a range of domain elements and asked to choose three, such that two are similar and different from the third. This is known as the method of triadic elicitation (e.g. Caputi and Reddy, 1999).<sup>\*</sup> In order to demonstrate this technique, suppose we were trying to uncover an astronomer's understanding of the planets within our own solar system. We might present her with a set of planets, and she might choose Mercury and Venus as the two similar elements and Jupiter as different from the other two. The expert is then asked for her reason for differentiating these elements, and this dimension is known as a construct. In our example, 'size' might be a suitable construct that differentiates between the selected elements. The remaining elements are then rated with respect to this construct.

This process continues with different triads of elements until the expert can think of no further discriminating constructs. The result is a matrix of similarity ratings, relating elements and constructs. This can be analysed using a variety of statistical techniques, of which the most popular is probably cluster analysis. Cluster analysis can reveal clusters of concepts, some of which may not have been articulated using other kinds of techniques (e.g. interviews).

Figure 7.1 shows the results of a repertory grid applied to the domain of planets (within our own solar system). We can see that the expert has so far generated seven constructs along which the planets vary. In this case, a nine-point rating scale has been used, and in the case of the 'size' (small/large) construct, the smallest planet, Mercury, has been given a rating of 1 and the largest planet, Jupiter, a rating of 9. The other planets have been rated in a comparative manner along the size construct.<sup>†</sup> The analysis has already revealed clusters of both constructs and elements.



**FIGURE 7.1** The results of the repertory grid technique applied to the domain of planets. (Implemented using WebGrid 5.)

<sup>\*</sup> In fact, Kelly (1955) describes a number of variations on the general triadic elicitation procedure. More information about these variations can be found in Fransella and Bannister (1977).

<sup>†</sup> In Figure 7.1, shading in the matrix is also used to highlight ratings. Heavy shading designates a high value for an element on a construct.



Thus, Jupiter and Saturn are clustered together at around 84% similarity, Neptune and Uranus at around 88% similarity, and these two pairs of clusters are themselves clustered together at around 79% similarity.\* An astronomer might well observe that this group of four planets constitutes the gas giants. A new concept – gas giant – has thus been uncovered, which might be distinguished from the other planets, that is the rocky or terrestrial planets. Note that Pluto bears very little similarity to other planets in the grid. In fact, it appears to occupy a category all by itself (although it does bear more similarity to the rocky planets than the gas giants). This is clearly interesting given the debate concerning the ontological status of Pluto as a proper planet.†

Constructs can also be the focus of cluster analysis. With respect to Figure 7.1, we can see that the constructs relating to temperature and distance from the Sun are clustered, as are the presence of rings and multiple moons. Such associations can reveal causal or other law-like relations in the domain, for example, the relationship between rings and moons may indicate some sort of causal relationship between the two.

Variants on the repertory grid technique allow you to run sociogrids (e.g. Shaw, 1980). These allow you to compare one individual's view of the domain with another's, and this can be important in terms of highlighting areas of consensus and difference among experts.

## LADDERED GRIDS

Another somewhat contrived technique that you will need to explain carefully to the expert before starting is the ladder grid technique. As part of this technique, the expert and elicitor construct a graphical representation of the domain in terms of the relations between domain and problem-solving elements. The result is a two-dimensional, hierarchically structured graph where nodes are connected by labelled arcs. No extra elicitation method is used here; expert and elicitor construct the graph together by negotiation.

In using the technique, the elicitor enters the conceptual map of the domain (see 'Concept Mapping and Process Mapping') at some point and then attempts to move around it with the expert. A formal specification of how we use the technique is shown here together with an example of its use.

Start the expert off with a seed item.

Move around the domain map using the following prompts:

- To move DOWN the expert's domain knowledge  
Can you give examples of <ITEM>?
- To move ACROSS the expert's domain knowledge  
What alternative examples of <CLASS> are there to <ITEM>?
- To move UP the expert's domain knowledge  
What have <SAME LEVEL ITEMS> got in common?  
What are <SAME LEVEL ITEMS> examples of?
- To elicit essential properties of an item  
How can you tell it is <ITEM> ?
- To discriminate items  
What is the key difference between <ITEM 1> and <ITEM 2>?

\* The similarity ratings between the individual elements and element clusters, in this case, are based on the FOCUS algorithm described by Jankowicz and Thomas (1982). The percentage similarity between adjacent elements in the grid is computed as  $((-100 * d)/c(n - 1)) + 100$ , where  $d$  is the sum of the absolute differences between the ratings of adjacent elements,  $c$  is the number of constructs in the grid (i.e. 7) and  $n$  is the largest rating possible (i.e. 9).

† See <http://news.bbc.co.uk/1/hi/5282440.stm>.

The elicitor may move around the knowledge map in any order which seems appropriate or convenient. As the session progresses, the elicitor keeps track of the elicited knowledge by drawing up a network on a large piece of paper or, if computer supported, via some other graphical characterisation. This representation allows the elicitor to make decisions (or ask questions) about what constitutes higher- or lower-order elements in the domain and what differences exist between elements in the network. In order to give the reader a flavour of the technique, there follows an extract from a laddered grid elicitation session. Once again, the knowledge domain is geology.

*KE:* So how could you tell something was dacite?

*EX:* Well + examine the fresh surface and the weathered surfaces first + looking at grain size, the relationship between the grains

*KE:* Can I just stop you there. What type of grain size is it?

*EX:* Coarse, medium, fine grain, oh, you want me to actually say what dacite is?

*KE:* The grain, in dacite what would it be?

*EX:* Er + medium grained.

*KE:* Medium grained, right. So can you give me other examples of medium grained rocks?

*EX:* Medium grained rocks + dolerite... Granodiorite as well... And we'll stay with that.

*KE:* Right, erm, what alternative is there to a medium grained rock?

*EX:* Well, you can have a coarse grained one or a fine grained one, those are sort of the three major ones.

*KE:* Right, can you give me examples of coarse grained rocks?

*EX:* Er, gabbro, granite... hmm, yeah, those two.

*KE:* And any examples of fine-grained rocks?

*EX:* Er, basalt... er andesite, trachyte... microgranite as well.

*KE:* Right, erm so. What about others

*EX:* Some of these are sort of a metamorphic ones where you're going to get large grains in a fine-grained matrix. There are phenocrysts in them, that's what we call the large grains

*KE:* Is, is there a word for that kind of texture or?

*EX:* Porphyritic mixture

*KE:* Can you give me the examples of the porphyritics...

*EX:* Nepheline-syenite, oh and Kentallenite

*KE:* How would you go about telling the difference between dolerite and granodiorite? What is the key difference?

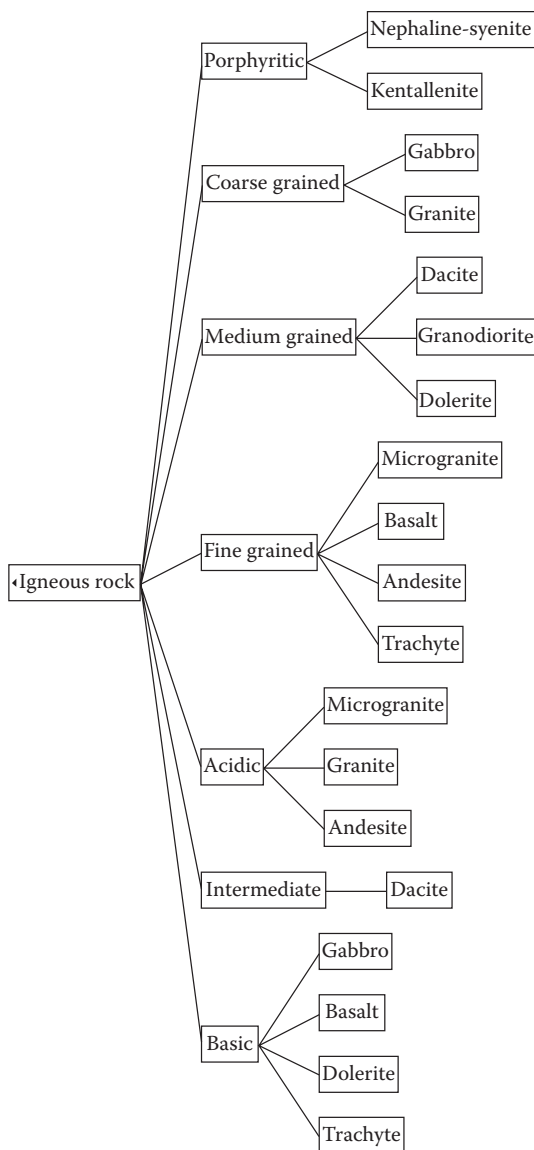
*EX:* Whether it's got quartz or hasn't got quartz or the percentage of quartz present will define whether it's an acidic rock or a basic rock, basic not having any quartz in it at all, and then er if there's a low amount, that's going to be an intermediate rock

*KE:* Which, which are the intermediate?

*EX:* Dacite + you've got high quartz are granite, microgranite, and andesite, and no quartz gabbro, basalt, dolerite and trachyte, intermediate dacite.

In the course of this laddered grid interview, the elicitor drew up a hierarchical representation of the domain as shown in Figure 7.2. This is only one of a number of representations that could have been made. In this case, the concepts of fine-, medium- and coarse-grained rocks have been understood to be classes of rock type. Similarly, the concept of an acidic, intermediate or basic rock has been treated as a class of rock type. However, the grain size and acidity (amount of quartz) could have been represented as properties of the particular rock types.

This hierarchy gives rise to the following set of rules that could be included in the knowledge base of a knowledge-intensive system for geological rock classification.



**FIGURE 7.2** Example of a ladder grid in the geology domain (this grid was developed using the Ladder Tool that is available as part of the PCPACK knowledge editing toolkit).

IF the rock is of medium grain size  
AND the rock is intermediate  
THEN the rock may be dacite

IF the rock is of coarse grain size  
AND the rock is acidic  
THEN the rock may be granite

IF the rock is of coarse grain size  
AND the rock is basic  
THEN the rock may be gabbro

As is the case with many knowledge elicitation techniques, it helps to keep an audio record of the session for future review or transcription. Laddering is an excellent way of carrying out a structured

interview. In addition, it is a technique that can be applied to a variety of knowledge types besides concepts, for example actions, tasks, goals, resources and so on can be the subject of a ladder grid knowledge elicitation session.

### LIMITED-INFORMATION TASK

A technique which can prove an excellent complement to the methods already outlined is a technique called the *limited-information task* (Hoffman, 1987) or *20 questions* (Grover, 1983). Using this technique, the expert is provided with little or no information about a particular problem to be solved, and the expert must therefore ask the elicitor for specific information that will be required to solve the problem. The information that is requested, along with the order in which it is requested, provides the elicitor with an insight into the expert's problem-solving strategy. One difficulty with this method is that the elicitor needs a good understanding of the domain in order to make sense of the expert's questions and to provide meaningful responses. The elicitor should have forearmed themselves with a problem from the domain together with a crib sheet of appropriate responses to the questions.

In one of the versions of the limited-information task that we use, we tell the expert that the elicitor has a scenario in mind, and the expert must determine what it is. The scenario might represent a problem, a solution or a problem context. The expert is told that they may ask the elicitor for more information, though what the elicitor gives back is terse (e.g. it may consist of simple 'yes' or 'no' responses) and does not go much beyond what was asked for in the question. The expert may be asked to explain why each of the questions was asked.

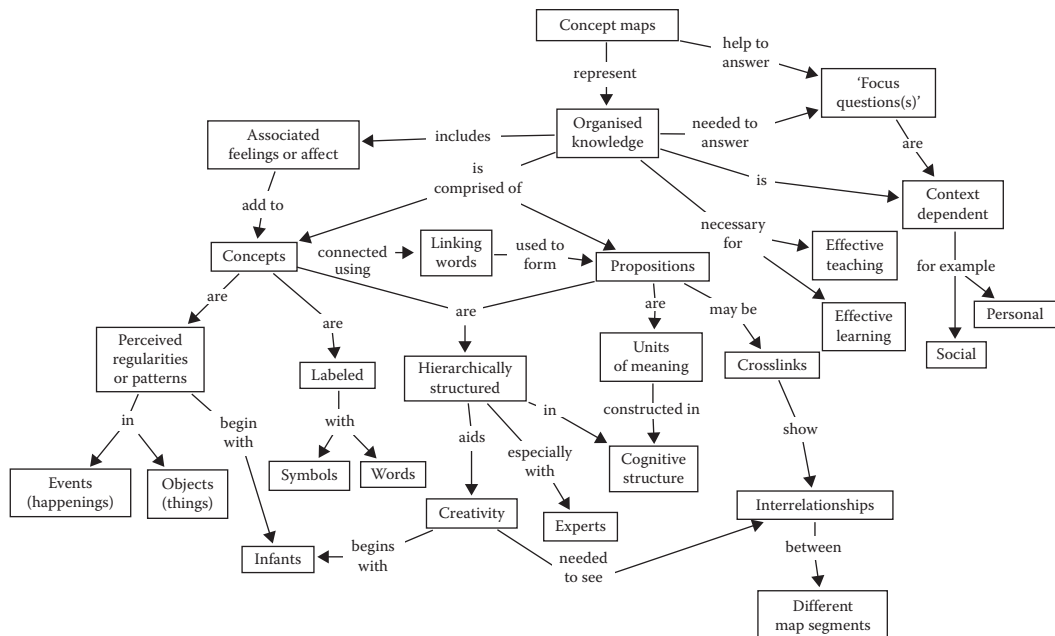
The limited-information task is useful because it provides information about the relative importance of particular items of information as part of a problem-solving process. Often, traditional knowledge-based systems gather the right data, but the order in which the data are gathered and used can be very different from how an expert works. This can decrease the acceptability of any implemented system if other experts are to use it, and it also has consequences for the intelligibility of any explanations the system offers in terms of a retrace of its steps to a solution.

The drawbacks to this technique are that the elicitor needs to have constructed plausible scenarios, and the elicitor has to be able to cope with the questions that are asked. The experts themselves are sometimes uncomfortable with this technique; this may well have to do with the fact that, as with other contrived techniques, it is not a natural means of manifesting expertise. In addition, whilst a few scenarios may reveal some of the general rules in a domain, the elicitation is very case specific. In order to get a broad range of knowledge, many different scenarios need to be constructed and used.

An interesting variation of this method is a form of telephone consultancy. Here we take two domain experts and place them at opposite ends of a table and ask them to imagine that one is a 'client' who is ringing up the other, a 'consultant', to ask for advice concerning a particular problem. They then engage in a conversation in which the 'consultant' tries to elicit the nature and context of the problem, and finally attempts to offer appropriate advice. In this variation of the limited-information task, you can rely on one of the experts to generate interesting cases. In addition, the expert playing the role of the 'client' can provide appropriate responses to the 'consultant's' enquiries. The only drawback is that sometimes experts construct extremely difficult cases for each other in order to test each other's mettle!

### CONCEPT MAPPING AND PROCESS MAPPING

Concept mapping and process mapping are both examples of diagramming techniques (Milton, 2012) that focus on the structure of conceptual and procedural knowledge, respectively. Concept mapping is probably one of the most widely used knowledge elicitation techniques, in part due to the popularity of the CmapTools software that was developed by the Institute for Human and Machine Cognition (IHMC) (see Knowledge Elicitation Tools).



**FIGURE 7.3** A concept map intended to explore the notion of a 'concept map'. (From <http://cmap.ihmc.us>.)

The artefacts that result from concept mapping (i.e. concept maps) are collections of propositions that are commonly displayed as a two-dimensional network of labelled grids and nodes (see Figure 7.3). Concept mapping has been reported to be a very efficient knowledge elicitation technique with the technique yielding an average of two useful propositions per session minute (Hoffman et al., 2001). The technique has also demonstrated its utility in a variety of disparate domains, with meteorology (Hoffman and Lintern, 2006) and intelligence analysis (Derbentseva and Mandel, 2011) serving as just a couple of examples.

Both concept and process mapping can be performed with popular knowledge acquisition toolkits, such as PCPACK and CmapTools (see Knowledge Elicitation Tools). In practice, however, CmapTools tends to be used primarily for concept mapping, while the features of the Diagram Tool within PCPACK make it ideally suited for process mapping. One of the features of the PCPACK Diagram Tool is a capability to 'drill down' into a process, detailing the structure of its constituent sub-processes. It also provides a range of process-oriented graphical notations that are consistent with those seen in popular modelling paradigms (e.g. UML activity diagrams).

## CLASSIFICATION OF KNOWLEDGE ELICITATION TECHNIQUES

We have now sampled some of the major approaches to knowledge elicitation and, where appropriate, given a detailed description of techniques that are likely to be of use. There are many variants on the methods we have described. We have provided a taxonomy of methods with which we are familiar together with a primary reference for each one as follows:

- Non-contrived/natural
  - Interviews
    - Structured
      - Fixed probe (Shadbolt and Burton, 1990a; Wood and Ford, 1993)
      - Focused interviews (Hart, 1986; Scott et al., 1991)
      - Forward scenario simulation (Grover, 1983)
      - Critical decision method (Hoffman et al., 1998)

- Semi-structured
  - Knowledge acquisition grid (LaFrance, 1987)
  - Teach back (Johnson and Johnson, 1987)
- Unstructured (Weiss and Kulikowski, 1984)
- Protocol analysis
  - Verbal
    - Online (Johnson et al., 1987)
    - Offline (Elstein et al., 1978)
    - Shadowing (Clarke, 1987)
    - Collegial verbalisation (Erlandsson and Jansson, 2007)
  - Behavioural (Ericsson and Simon, 1996)
- Contrived
  - Diagramming
    - Laddered grid (Corbridge et al., 1994; Walker and Crittenden, 2012)
    - Concept mapping (Novak and Cañas, 2006)
    - Process mapping (Milton, 2012)
  - Sorting and rating
    - Concept sorting (Gammack, 1987)
    - Repertory grid (Shaw and Gaines, 1987)
    - Pathfinder (Schvaneveldt et al., 1985)
  - Constrained processing
    - Limited-information task (Hoffman, 1987)
    - 20 Questions (Grover, 1983)

This is, of course, only one possible structure for a taxonomy of knowledge elicitation techniques. A number of alternative classifications appear in the literature based on a variety of perspectives, such as the nature of the interaction between elicitor and expert, the type of knowledge (conceptual vs. procedural) elicited from the expert and the kind of materials required by the task or delivered as outputs from the task. Gavrilova and Andreeva (2012) categorise knowledge elicitation based on the level of involvement of an expert and an elicitor and type of interaction/collaboration between them. They distinguish between ‘active’ (analyst-leading) and ‘passive’ (expert-leading) techniques, where an active technique requires ‘the active position of an analyst, who “pulls” the knowledge from the expert with the help of specially prepared questions’ and a passive technique is a technique in which ‘the analyst’s interference into the process in which the expert is engaged is very limited’ (Gavrilova and Andreeva, 2012, p. 529). As another example of the taxonomic organisation of knowledge elicitation techniques, Milton (2012) organises knowledge elicitation techniques into three categories, namely questioning techniques (e.g. laddering), task-based techniques (e.g. concept sorting) and diagramming techniques (e.g. concept mapping).

None of the existing taxonomies (including the one presented here) are necessarily complete with respect to the range of knowledge elicitation techniques that have been discussed in the literature. In part, this stems from the fact that the goals of knowledge elicitation and the kind of task contexts in which knowledge elicitation is deemed important have changed over time. As pointed out by Hoffman and Lintern (2006), the methodology of knowledge elicitation could be folded into the broader methodology of cognitive task analysis, which is a focal point for human factors and cognitive systems engineering. This serves to blur the distinction between knowledge engineering and cognitive engineering, and it tends to result in a greatly expanded palette of knowledge elicitation methods. A variety of ethnographic methods, for example, could be seen as forms of knowledge elicitation (see Hutchins, 1995).

Other techniques that are sometimes presented as knowledge elicitation techniques are the various methods associated with data mining (Witten and Eibe, 2005), machine learning (Mitchell, 1997) and rule induction (Hart, 1986). These techniques are not covered in

detail here because they are not techniques that are typically used in conjunction with domain experts. There are, however, some exceptions. In particular, there have been a number of recent attempts to combine expert input with machine learning techniques in order to improve the quality of the knowledge that results from the machine learning process. Typically, the kind of outputs delivered by machine learning tend to prove difficult for experts to understand and extend, and this presents problems in terms of the maintenance of the knowledge base and the trust that experts place in automated decision-making processes. Argument-based machine learning (ABML) is a technique which was developed to address some of these issues (Mozina et al., 2008). The technique is intended to combine expert knowledge with machine learning processes, and it requires the expert to explain the reasons for decisions in particular cases. Groznik et al. (2013) describe a recent application of the technique, wherein ABML is used to elicit knowledge from neurologists in order to develop a decision support system concerned with neurological diagnoses.

## EXPERTS AND EXPERTISE

As the source of much of the knowledge that is captured as part of a knowledge engineering initiative, domain experts are a critical focus of attention for those involved in knowledge elicitation. Failing to pay adequate attention to the differences among experts, as well as the level of expertise they possess, is likely to have a profound effect on the efficiency of the knowledge elicitation process, as well as the quality of the knowledge that gets elicited.

One of the first challenges that must be addressed in any knowledge engineering project is the identification of individuals with the relevant expertise. In some cases, it may be obvious who the experts are within a given domain; in other cases, however, it may not at all be clear how experts should be identified. Factors such as the possession of professional qualifications, experience and occupational position, as well as the results of testing and screening processes, may all be used as the basis for expert identification; however, none of these methods is without its problems (Farrington-Darby and Wilson, 2006). For example, the position held by an individual is a commonly used criterion for expert selection; however, the reasons for individuals being awarded a position within a given occupational setting may have very little to do with their actual expertise (see Farrington-Darby and Wilson, 2006). In terms of experience, a general rule of thumb is that expertise develops after about 10,000 h of practice. Recent research, however, has suggested that expertise in some domains, such as weather forecasting, may take considerably longer (see Hoffman and Lintern, 2006). In spite of the difficulties, it is worth spending some time considering who and who is not an expert. As Burton et al. (1990) note

Inadequate expertise is likely to continue to be a problem for those working in applied settings. We suggest that considerable time be put into the original selection of an expert. External validation of an expert's suitability will save considerable time and wasted effort in future sessions. (p. 177)

Once experts have been identified, it is important to consider the differences between experts, as well as the nature of the expertise they manifest. Experts can be differentiated in a number of ways; however, one scheme that we have found useful in practice is to distinguish between three kinds of experts: the *academic*, the *practitioner* and the *samurai*. Each of these types of expert differs along a number of dimensions.\* These include the outcome of their expert deliberations, the problem-solving environment they work in, the state of the knowledge they possess (both its internal structure and its external manifestation), their status and responsibilities, their source of information and the nature of their training.

---

\* In practice, of course, experts do not tend to fall in one or other categories; rather, they embody elements of all three types of expert.



How are we to tell these different types of expert apart when we encounter them? The academic type regards their domain as having a logically organised structure. Generalisations over the laws and behaviour of the domain are important to them; theoretical understanding is prized. Part of the function of such experts may be to explicate, clarify and teach others. They thus talk a lot about their domains. They may feel an obligation to present a consistent story for both pedagogic and professional reasons. Their knowledge is likely to be well structured and accessible. These experts may suppose that the outcome of their deliberations should be the correct solution of a problem. They believe that the problem can be solved by the appropriate application of theory. They may, however, be remote from everyday problem solving.

The practitioner class, on the other hand, are engaged in constant day-to-day problem solving in their domain. For them, specific problems and events are the reality. Their practice may often be implicit, and what they desire as an outcome is a decision that works within the constraints and resource limitations in which they are working. It may be that the generalised theory of the academic is poorly represented and articulated by the practitioner. For the practitioner, heuristics may dominate and theory is sometimes thin on the ground.

The samurai is a pure performance expert – their only reality is the performance of action to secure an optimal performance. Practice is often the only training, and responses are often automatic.

One can see this sort of distinction between experts in any complex domain. Consider, for example, medical domains where we have professors of the subject, busy doctors working the wards and medical ancillary staff performing many important but repetitive clinical activities.

The knowledge elicitor must be alert to these differences because the various types of expert will perform very differently in knowledge elicitation situations. The academic will be concerned to demonstrate mastery of the theory. They will devote much effort to characterising the scope and limitations of the domain theory. Practitioners, on the other hand, are driven by the cases they are solving from day to day. They have often compiled or routinised any declarative descriptions of the theory that supposedly underlies their problem solving. The performance samurai will more often than not turn any knowledge elicitation interaction into a concrete performance of the task, simply exhibiting their skill.

Another important distinction between experts is with respect to their level of expertise. A number of models of expertise development have been proposed within the cognitive science and human factors communities, and these may serve as the basis for a second dimension along which experts can be classified – one that is largely orthogonal to the previously mentioned distinction between academics, practitioners and samurai. One model, proposed by Dreyfus and Dreyfus (1986), suggests that expertise develops via the progression through five sequential stages: novice, advanced beginner, competent, proficient and expert. The transition between these stages is assumed to depend on the accumulation of situated practical experience within the relevant domain. Another classification scheme derives from the Craft Guilds of the Middle Ages (Hoffman, 1998; Hoffman et al., 1995). In this case, the developmental scale ranges from a 'naivette' (i.e. one who is totally ignorant of a domain) through to a 'master' who is regarded as one of an elite group of experts – the expert of experts.

Recognising the developmental stage of an expert can be important for the purposes of knowledge elicitation. Clearly, individuals with well-developed levels of expertise are important targets for knowledge elicitation, since they are the ones who are likely to possess the greatest amount of domain-relevant knowledge. Having said that, expertise development tends to be associated with a shift from explicit to tacit knowledge, and thus individuals at different points on the developmental trajectory from novice to master may be differentially responsive to particular kinds of knowledge elicitation technique. In certain kinds of domains, for example, a 'journeyman' or 'expert' may have greater conscious access to domain-relevant knowledge as compared to a 'master'. For this reason, techniques such as interviews may yield more information from those at intermediate levels of expertise development as compared to those further along the developmental scale.

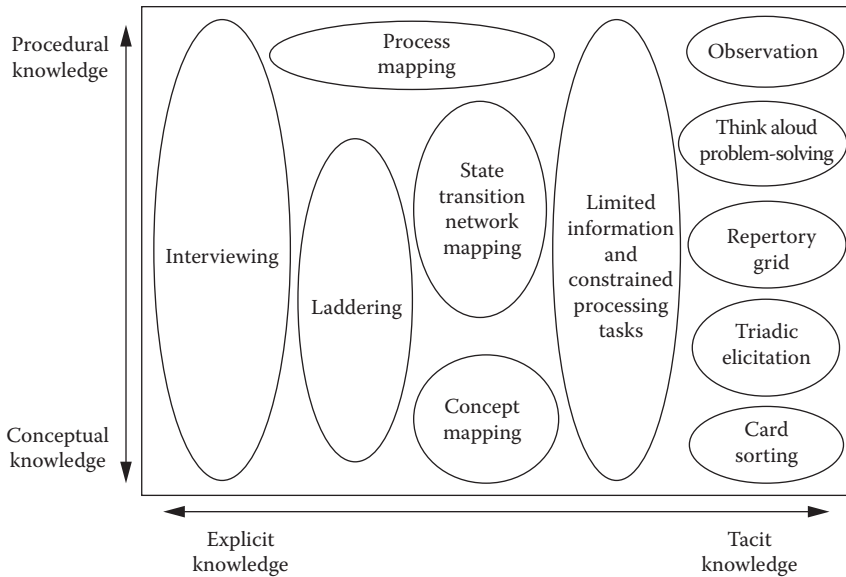
Clearly, the expertise embodied by experts is not of a homogenous type (Feltovich et al., 1997). In constructing any knowledge-intensive system, it is likely that very different types of knowledge will be uncovered, and these are likely to have very different roles in the system under development. In general, we can distinguish between four kinds of knowledge (three of these – the domain, inference and task knowledge categories – are explicitly represented within knowledge engineering methodologies, such as the CommonKADS methodology (see Schreiber et al., 2000)):

- *Domain knowledge.* First, we can distinguish what is called domain knowledge. This term is being used in the narrow sense of knowledge that describes the concepts and elements in the domain and relations between them. This sort of knowledge is sometimes referred to as declarative knowledge – it describes what is known about things in the domain.
- *Inference knowledge.* There is also knowledge and expertise that has to do with what we might call the inference level. This is knowledge about how the components of expertise are to be organised and used in the overall system. It tells us the type of inferences that will be made and what role knowledge will play in those inferences. This is quite a high-level description of expert behaviour and may often be implicit in expert practice.
- *Task knowledge.* Another type of expert knowledge is the task level. This is sometimes called procedural knowledge. This is knowledge concerned with goals, sub-goals, tasks and sub-tasks. Thus, in a classification task, there may exist a number of tasks to perform in a particular order so as to utilise the domain-level knowledge appropriately.
- *Strategic knowledge.* Finally, there is a level of expert knowledge referred to as strategic knowledge. This is information that monitors and controls the overall problem-solving process.

Within any of these categories of knowledge, the information may be either implicit or explicit. Thus, in some domains, the expert may have no real notion of the strategic knowledge they are following, whilst in others, this knowledge is very much at the forefront of their deliberations.

## METHODOLOGIES AND PROGRAMMES

We turn next to the question of how knowledge elicitation techniques should be assembled to form a programme of knowledge acquisition. There are a number of articles and books on how to undertake knowledge elicitation as part of a knowledge engineering project. Milton (2007), for example, describes the processes involved in knowledge elicitation and modelling in the form of a step-by-step guide. The choice as to which knowledge elicitation technique to use in any particular situation is guided by a variety of criteria, including the characteristics of the domain, the nature of the domain expert and the requirements associated with the proposed knowledge system solution. Furthermore, it is clear that some techniques are going to be more costly in terms of time with the expert or else the effort required for the analysis of elicited material. In order to select an appropriate knowledge elicitation technique, one needs to understand which method best fits the particular problem and situation. This calls for empirical evaluations of each of the techniques with respect to factors such as the nature of experts and their associated expertise. Although there are a variety of difficulties associated with the evaluation of knowledge elicitation techniques (Shadbolt et al., 1999), the available research has provided some general conclusions as to their relative efficacy (Burton et al., 1987, 1990; Hoffman et al., 1995; Shadbolt and Burton, 1990b). It has also provided some guidelines as to when to use particular kinds of knowledge elicitation technique. Gammack and Young (1985), for example, offer a mapping of knowledge techniques onto domain types. Their analysis requires that domain knowledge be separated into



**FIGURE 7.4** Differential utility of knowledge elicitation techniques with respect to the elicitation of different kinds of knowledge. (From Milton, N., *Personal Knowledge Techniques*, University of Nottingham, Nottingham, U.K., 2003.)

different categories, and they provide suggestions about which techniques are most likely to be effective within each category.

One of the main criteria for choosing between different techniques within a programme of knowledge elicitation is likely to be the type of knowledge that needs to be elicited. In this respect, the distinction between explicit and tacit knowledge has proven to be of significant interest. Different knowledge elicitation techniques are thus deemed to be differentially effective at eliciting explicit or tacit knowledge (see Figure 7.4). Another knowledge dimension that is often seen as important is the distinction between conceptual and procedural knowledge. Here, techniques such as process mapping are considered to be more effective for the elicitation of procedural knowledge, and techniques such as concept mapping and concept sorting are deemed to be more effective for the elicitation of conceptual knowledge. Figure 7.4 summarises the differential suitability of a number of knowledge elicitation techniques with respect to these two knowledge dimensions (i.e. explicit/implicit and conceptual/procedural).

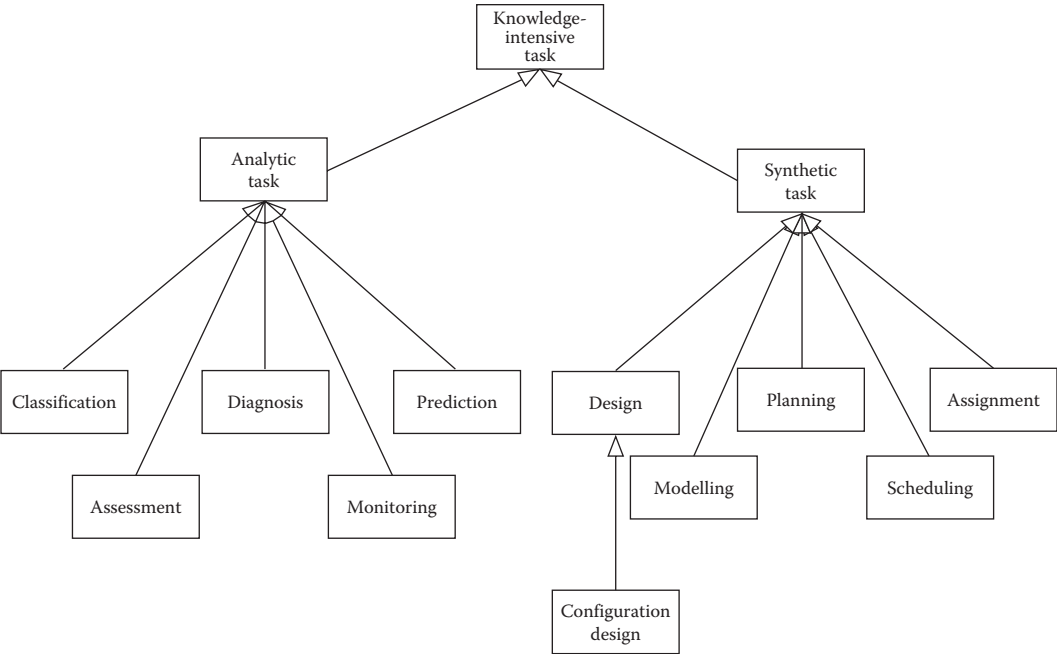
The notion that different knowledge elicitation methods are differentially effective at eliciting particular kinds of knowledge forms part of what has become known as the differential access hypothesis (Hoffman et al., 1995). Although some empirical support for the hypothesis has been found, a strong version of the differential access hypothesis (namely the idea that certain kinds of knowledge can be elicited *only* via the use of particular techniques) remains a point of contention within the knowledge engineering community (Hoffman and Lintern, 2006). When it comes to the notion of tacit knowledge, for example, Hoffman and Lintern (2006) suggest that the different knowledge elicitation techniques establish different conditions under which the verbalisation of tacit knowledge is more or less likely. They suggest knowledge elicitation techniques should be seen as ‘scaffolds’ that support the expression or communication of knowledge. With this in mind, the key aim in knowledge elicitation becomes one of establishing the right kind of conditions under which experts can articulate, or otherwise communicate, their expertise. These kinds of conditions are clearly influenced by the kind of technique that is used, since each technique is associated with different forms of social interaction,

access to mnemonic cues, the use of different diagrammatic representations and so on. With this in mind, it might be argued that something like tacit knowledge should not be seen as a form of knowledge that can never, in principle, be verbalised by experts; rather, it should be seen as a form of knowledge that is more easily articulated in certain situations as opposed to others. This, suggest Hoffman and Lintern (2006), has shifted the debate from a consideration of differential access to one of differential utility when it comes to the selection of knowledge elicitation techniques:

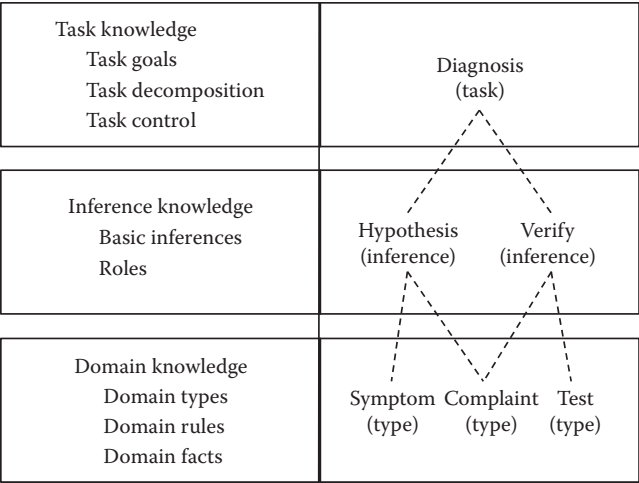
The hypothetical problem of differential access has given way to a practical consideration of differential utility. Any given method might be more useful for certain purposes, might be more applicable to certain domains, or might be more useful with certain experts having certain cognitive styles. In other words, each knowledge elicitation method has its strengths and weaknesses. Some of these are purely methodological or procedural (e.g. transcription and protocol analysis takes a long time), but some relate to the content of what is elicited. (Hoffman and Lintern, 2006, pp. 216–217)

In spite of this change in perspective, however, it should be clear that there remains a compelling reason to exploit a variety of techniques within any programme of knowledge elicitation. Even when it appears that only one particular body of knowledge is being dealt with – one which shows no internal differentiation with respect to, for example, explicit/tacit or procedural/conceptual distinctions – it is still advisable to use a variety of techniques. One reason for this stems from the possibility that the knowledge elicited by different techniques may predict actual performance to a greater or lesser extent. Studies have thus found that the content of verbal reports and the details of actual performance are not always the same. Cooke and Breedin (1994), for example, discovered a dissociation between the written explanations that were offered for physics trajectory problems and the actual predictions that were made concerning those trajectories. These results suggest that the results of multiple techniques should be compared with each other in order to evaluate the connection between knowledge and performance.

One of the factors that may inform the design of knowledge elicitation programmes is the methodological framework in which knowledge elicitation and modelling are undertaken. Although a number of methodologies exist for the development of ontologies within the context of the Semantic Web (e.g. Sure et al., 2003), such methodologies typically ignore the early steps of the knowledge engineering process and place little emphasis on knowledge elicitation. CommonKADS (Schreiber et al., 2000) is one of the few methodologies that explicitly incorporates the use of knowledge elicitation techniques. One way in which CommonKADS helps to structure the knowledge elicitation activity is by the distinction it makes between domain, task and inference knowledge (see Experts and Expertise). These different kinds of knowledge are represented as distinct ‘layers’ within a CommonKADS knowledge model specification, and mappings are established between the layers (e.g. between elements of inference and domain knowledge) in order to flexibly link different kinds of knowledge together in the context of a particular knowledge solution (see Figure 7.6). CommonKADS also offers a range of reusable components that can be used as points of departure for the selection and implementation of knowledge elicitation activities. The reusable components include a set of domain schemas, a catalogue of inference types and a library of task templates. These are useful not only in terms of improving the efficiency of the modelling process; they also serve to focus attention on the kinds of knowledge that needs to be acquired in the context of a particular kind of knowledge-based activity. Each of the CommonKADS task templates (see Figure 7.5) thus highlights the typical pattern of inferences that are associated with each kind of task, and it also links these inferences with particular bodies of domain knowledge (e.g. concepts) (see Figure 7.6). This kind of information can be extremely valuable in terms of highlighting the kind of knowledge to elicit and the kind of behavioural patterns to look for in expert performance.



**FIGURE 7.5** Knowledge-intensive tasks recognised by the CommonKADS methodology. Each of these tasks is associated with default inferences, control structures and template domain schemas.



**FIGURE 7.6** Linkages between the various layers of the CommonKADS knowledge model for a particular kind of knowledge-intensive task – in this case, diagnosis. Each task is associated with specific types of inferences that are themselves linked with particular elements at the level of domain knowledge.

**KNOWLEDGE ELICITATION TOOLS**

As indicated in the previous section, the attempt to improve our understanding of the conditions under which knowledge elicitation techniques are most effective, as well as how to adapt those techniques within specific knowledge elicitation programs, is the focus of recent and ongoing research attention. Another focus of attention concerns the development of software tools to support the knowledge elicitation process.

The software tools that are presented in this section – PCPACK, Protégé and CmapTools – have a long history of development and use within the knowledge acquisition community. The recent development of these tools has been strongly influenced by the web\* and, in particular, the Semantic Web. All the tools have thus been extended in particular ways to accommodate the representational frameworks associated with the Semantic Web. Recent versions of PCPACK thus provide support for RDF export, while knowledge elicitation plug-ins for Protégé interoperate with the Protégé-OWL plug-in in order to provide support for knowledge elicitation in the context of ontology development (Wang et al., 2006). There has also been a recent effort to extend CmapTools in order to provide support for the visualisation and editing of OWL ontologies (Eskridge and Hoffman, 2012; Hayes et al., 2005).

## PCPACK

PCPACK is an integrated suite of knowledge elicitation tools that has a long history of use within the knowledge engineering community (Schreiber et al., 2000, Chapter 8). Early versions of PCPACK provided computerised support for many of the knowledge elicitation techniques described earlier in this chapter (O'Hara et al., 1998; Shadbolt and Milton, 1999); however, more recent versions of the software have settled on those tools that provide the greatest level of support to those engaged in corporate knowledge engineering and management initiatives. The current version of PCPACK is maintained and distributed by Tacit Connexions, and a fully operational demonstration version of the software can be downloaded from the Tacit Connexions website.<sup>†</sup> PCPACK includes a variety of tools to support knowledge elicitation and modelling, and all of these tools are integrated with a single knowledge repository such that any changes to the knowledge base made using one tool are immediately reflected in other components of the tool suite. Among the tools included with PCPACK is the Ladder Tool, which is used for creating ladder grids of various kinds (e.g. taxonomic and meronymic concept hierarchies); a Diagram Tool, which can be used for process and concept mapping; a Protocol Tool, which can be used for PA; an Annotation Tool, which is used to provide an HTML editing interface for knowledge objects; and a Publisher Tool, which enables knowledge models to be published as web-accessible 'Knowledge webs'. Other tools provide support for RDF import/export, annotation template management and the matrix-based editing of knowledge object properties and relationships. Figure 7.7 shows a screenshot of one of the PCPACK tools, namely the Ladder Tool.

## Protégé

As with PCPACK, the Protégé knowledge editor<sup>‡</sup> has a long history of use within the knowledge engineering community. As a flexible and customisable knowledge editing environment, Protégé is able to provide support for a variety of knowledge engineering methodologies and modelling frameworks. However, ever since the advent of the Semantic Web and the development of the Protégé-OWL plug-in (Knublauch et al., 2004, 2005), it is probably fair to say that the primary use of the tool is to develop (OWL-based) ontologies.

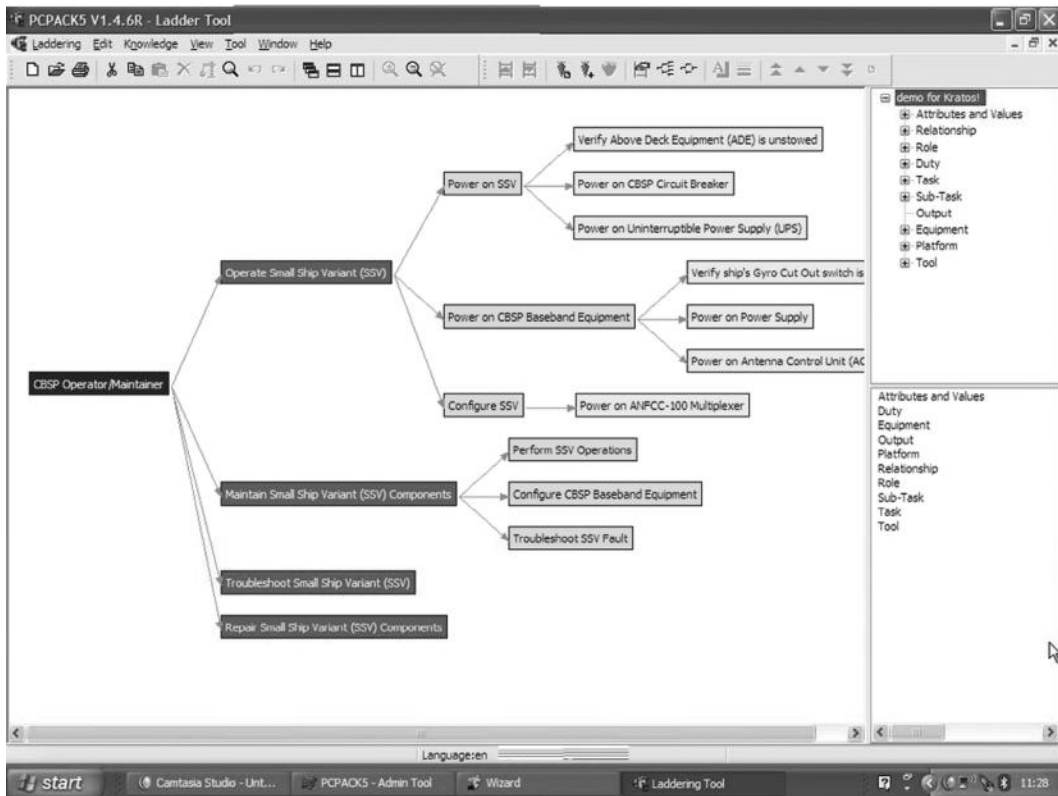
Unlike PCPACK, Protégé does not provide an integrated suite of knowledge elicitation tools as standard. The primary purpose of the tool is to support the editing of elicited knowledge rather than to support the process of knowledge elicitation itself. There have, however, been a number of attempts to provide computerised versions of the knowledge elicitation techniques as plug-ins to the Protégé environment. Wang et al. (2006) thus describe the attempt to implement card-sorting

\* For example, all the tools reviewed in this section support the publication of HTML versions of knowledge models. This enables the models to be accessed in the context of the conventional, document-centred web, as well as the more recent data-centric web or web of linked open data (see Heath and Bizer, 2011).

<sup>†</sup> See <http://www.tacitconnexions.com/PCPACK%20download%20promo%20page.htm>.

<sup>‡</sup> See <http://protege.stanford.edu/>.





**FIGURE 7.7** The PCPACK ladder tool.

and laddering plug-ins in order to support the use of knowledge elicitation techniques as part of the ontology development process.

Protégé is available as a free, open-source download from the Protégé website. It has typically been implemented as a Java-based desktop application; however, recent development efforts have seen the release of WebProtégé (Tudorache et al., 2013), which is a lightweight, web-based version of the original Protégé environment.

## CmapTools

Another widely used knowledge elicitation and knowledge modelling tool is CmapTools, which is developed and maintained by the IHMC.\* CmapTools provides support for the development of concept maps, which can be developed in conjunction with a domain expert and then published on the web. The tool enables the user to establish links between concept maps, which are collectively referred to as a 'knowledge model'. In addition, links to other resources, such as images, videos, text documents and so on, can be associated with any node in the concept map diagram.

As with other knowledge engineering technologies, the development of CmapTools is currently being influenced by the Semantic Web. Researchers at the IHMC are currently exploring the potential to combine concept mapping capabilities with the representational formalisms encountered in the context of the Semantic Web (Eskridge and Hoffman, 2012; Hayes et al., 2005). Ultimately, this effort will enable the CmapTools concept mapping system to be used for the construction, sharing and visualisation of OWL ontologies.

\* See <http://cmap.ihmc.us/>.



## KNOWLEDGE ELICITATION, KNOWLEDGE ENGINEERING AND THE WORLD WIDE WEB

The web and the Semantic Web have had a profound impact on the discipline of knowledge engineering (Gil, 2011; Schreiber, 2013). In many cases, the web now serves as both the starting point (e.g. by providing access to a rich source of domain-relevant knowledge and information) and the end point (e.g. by serving as a platform for knowledge publication and distribution) for knowledge engineering efforts. The specifications and recommendations that have emerged in the context of the Semantic Web initiative (Berners-Lee et al., 2001; Shadbolt et al., 2006) (e.g. RDF, RDF-S and OWL) have served as a Procrustean bed that has affected nearly all knowledge representation frameworks and knowledge engineering technologies. The web is also an environment that can be used for the purposes of knowledge elicitation, especially when the elicitation effort requires collaboration from multiple stakeholders. Finally, of course, the web serves as an environment for the implementation of a whole variety of intelligent systems and knowledge-based solutions.

Perhaps the most notable feature of the web, when it comes to knowledge elicitation, is the role that the web plays as a knowledge source. The web provides access to a rich range of resources that are relevant to the construction of any prospective knowledge-intensive system. If one takes the domain used throughout much of this chapter – the classification of rocks and minerals – one is able to find a wealth of online resources. These range from dictionaries and definitions of terms, succinct summaries of the processes of rock formation and extensive online databases. Such resources can serve as an important focus for the initial stages of knowledge elicitation, particularly for the purposes of domain familiarisation. They can also provide access to a range of materials that can be incorporated into knowledge elicitation exercises (e.g. images of different kinds of rocks can be used as the basis for card-sorting exercises).

Web-based resources may also serve as the direct target of knowledge acquisition efforts. Although such resources by themselves are unlikely to provide all the required information – recall the aphorism ‘the gold is not in the documents’ – they can yield knowledge structures (e.g. concept lists) that are subsequently refined and extended in the course of face-to-face knowledge elicitation sessions.

Complementing the use of manual knowledge acquisition methods is the use of a range of advanced knowledge discovery techniques that can be used to extract knowledge from online sources. These kind of automated techniques are vitally important given the scale of the web and the range of resources that are now available. Information extraction and natural language processing (NLP) technologies are one focus of ongoing research attention in this area (Sarawagi, 2008), as are opinion mining and sentiment analysis techniques (Feldman, 2013; Pang and Lee, 2008). There is also interest in the use of ontology learning techniques to create initial ontological structures from large-scale bodies of domain-relevant information (Maedche and Staab, 2003). These kinds of analytic and learning techniques are likely to become all the more important as we move into an era where linked open data assets (see Heath and Bizer, 2011) become increasingly prevalent on the web.

Web resources may also be used as part of an integrated knowledge acquisition effort that combines web access with the use of conventional knowledge elicitation techniques and other forms of advanced machine-based processing, such as NLP. Mendonça et al. (2012) thus used NLP to isolate initial concepts and then refined these in conjunction with domain experts using a variety of knowledge elicitation techniques (namely interviews, sorting and matrix-based techniques). This was followed by a knowledge validation phase in which the web was used to support the collaborative validation of elicited knowledge. This study highlights how the web can be exploited at several stages of the knowledge elicitation process: it can be used as an initial resource to support domain familiarisation and extract initial concepts (perhaps using machine-assisted techniques, such as NLP), and it can also be used to validate the elicited knowledge – the knowledge is published on the web and made available to a global community of experts who can validate and refine the elicited knowledge as a precursor to, for example, ontology development. Further research in this area

should consider the kind of opportunities the web makes available for knowledge elicitation and adapt knowledge engineering methodologies to exploit these opportunities.

The main problem, of course, when it comes to the use of web-based resources concerns their varying quality and coverage. The information provided by the sources is often of unknown origin, and there is often no prior history with many of the sources that may be used to assess their reputation. One focus of ongoing research is how to determine whether to trust a particular piece of information provided by a source.

In addition to the use of the web as a knowledge source, the web also provides a platform for active knowledge elicitation from individual experts or expert communities. Unfortunately, there are very few examples, at the present time, of web-based tools that could be used for collaborative knowledge elicitation. Perhaps one reason for this relates to a shift in our appreciation of how the web can be used as a mechanism for knowledge acquisition. When one looks at examples of large knowledge repositories on the web – for example Wikipedia – what one tends to encounter is a system in which knowledge content has emerged as a result of the collaborative efforts of multiple individuals. This has led to our traditional notions of expert-centred knowledge engineering being supplemented with an approach that draws on the contributions of large numbers of users, very few of whom are perhaps regarded as experts within the target domain. The point is that sometimes the actions of a large number of users can yield useful knowledge outputs (although whether these outputs can ever serve as a substitute for the kind of outputs obtained in face-to-face knowledge elicitation sessions with domain experts is currently a moot point). Folksonomies (Wu et al., 2006) represent one example here, as do the structured resources that emerge from the cumulative editing actions of Wikipedia users, for example DBpedia (Bizer et al., 2009). In general, there is an increasing recognition of the way in which certain classes of web-based systems – sometimes referred to as social machines – can be used to leverage the contributions of human user communities, often at large scale. Knowledge acquisition is often a key focus of such systems (Shadbolt, 2013); however, the systems can also (on occasion) yield collective problem-solving performances that parallel those of individual human experts. In such cases, it may be possible to see a social machine as a form of biotechnologically hybrid intelligent system that dynamically exploits the complementary contributions of both human individuals and conventional computing systems.

One final point that is worth reiterating here relates to the way in which the Semantic Web has impacted knowledge engineering efforts. As mentioned previously, many of the tools used for knowledge elicitation have been influenced by the advent of ontology languages that have been developed for the Semantic Web, and the output of many knowledge engineering efforts now consists in the generation of resources (e.g. OWL ontologies) that are compliant with the standards and recommendations of the Semantic Web community. It is tempting to think of the Semantic Web, in this case, as a large-scale knowledge repository that is the distributed counterpart of the more centralised knowledge bases encountered in the era of expert systems development. There are, however, a number of differences between the Semantic Web and conventional knowledge bases, of which the most obvious relate to the heterogeneity, scale and diverse quality of Semantic Web knowledge content (d'Aquin et al., 2008). It is also fair to say that the content of the Semantic Web tends to be used in a manner that is unlike that seen in the case of conventional expert systems. As Brueker (2013) notes, 'Ontologies are rarely used as knowledge bases, but rather as (shallow) vocabularies for managing large information repositories' (p. 179). Indeed, as is evidenced by systems such as IBM's Watson (Ferrucci et al., 2010), intelligence on the Semantic Web is likely to emerge as a result of the ability to exploit large amounts of available data rather than an ability to carry out sophisticated reasoning (d'Aquin et al., 2008). Although Watson does use ontologies for some inferences, its answers are, for the most part, based on sophisticated information retrieval capabilities and the ability to integrate probabilistic evidence from many diverse sources.

The ability to treat the web as an epistemic resource and press maximal benefit from an ever-expanding quantity of linked data assets is likely to be a key focus area for research into the next generation of intelligent systems. To what extent computational ontologies will play a role in the

realisation of these capabilities is unclear; however, what is largely beyond dispute is that, in the near future, the web is likely to serve as a means by which human knowledge is made available for a variety of purposes, and in view of this, the interest in knowledge elicitation and the need for robust knowledge elicitation techniques is likely to continue.

## CONCLUSION

Despite a range of scientific and technical advances (including the continued development of the web and Semantic Web), the problem of knowledge elicitation remains an important area of research attention and practical application. This chapter has described some of the methods and techniques that are used in this enterprise. We have also sought to provide an indication of the difficulties inherent in doing this kind of work. Knowledge elicitation is itself a form of complex expertise. Experienced knowledge engineers come to recognise the characteristics of expert thinking, and they develop skills that allow them to capture an expert's knowledge despite the many obstacles they face. Continued research into the differential effectiveness of knowledge elicitation techniques in different situations is likely to inform our understanding of how to structure and manage the knowledge acquisition process; however, there really is no substitute for real-world practical experience when it comes to knowledge elicitation. Just as expertise in other areas comes only at the expense of many hours of practical experience within the relevant domain, so a mastery of knowledge elicitation often requires many hours of active engagement in the knowledge elicitation process.

## REFERENCES

- Bainbridge, L. and Sanderson, P. (2005) Verbal protocol analysis. In J. Wilson, T. Megaw and N. Corlett (Eds.), *Evaluation of Human Work*. Taylor & Francis Group, London, U.K.
- Beilock, S. L., Wierenga, S. A. and Carr, T. H. (2002) Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *The Quarterly Journal of Experimental Psychology: Section A*, 55(4), 1211–1240.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001) The semantic web. *Scientific American*, 284(4), 34–43.
- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. and Hellmann, S. (2009) DBpedia – A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3), 154–165.
- Brueker, J. (2013) A cognitive science perspective on knowledge acquisition. *International Journal of Human-Computer Studies*, 71(2), 177–183.
- Burton, A. M., Shadbolt, N. R., Hedgecock, A. P. and Rugg, G. (1987) A formal evaluation of knowledge elicitation techniques for expert systems: Domain 1. In D. S. Moralee (Ed.), *Research and Development in Expert Systems IV*. Cambridge University Press, New York.
- Burton, A. M., Shadbolt, N. R., Rugg, G. and Hedgecock, A. P. (1990) The efficacy of knowledge elicitation techniques: A comparison across domains and levels of expertise. *Knowledge Acquisition*, 2(2), 167–178.
- Caputi, P. and Reddy, P. (1999) A comparison of triadic and dyadic methods of personal construct elicitation. *Journal of Constructivist Psychology*, 12(3), 253–264.
- Chi, M. T. H., Feltovich, P. J. and Glaser, R. (1981) Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152.
- Chi, M. T. H., Glaser, R. and Farr, M. J. (Eds.) (1988) *The Nature of Expertise*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Chi, M. T. H., Glaser, R. and Rees, E. (1982) Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence: 1*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Clarke, B. (1987) Knowledge acquisition for real time knowledge based systems. *First European Workshop on Knowledge Acquisition for Knowledge Based Systems*, Reading, U.K.
- Cooke, N. J. (1994) Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41(6), 801–849.
- Cooke, N. J. (1999) Knowledge elicitation. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay and M. T. H. Chi (Eds.), *Handbook of Applied Cognition* (pp. 479–510). John Wiley & Sons, Ltd, Chichester, U.K.

- Cooke, N. J. and Breedin, S. D. (1994) Constructing naive theories of motion on the fly. *Memory & Cognition*, 22(4), 474–493.
- Corbridge, C., Rugg, G., Major, N. P., Shadbolt, N. R. and Burton, A. M. (1994) Laddering: Technique and tool use in knowledge acquisition. *Knowledge Acquisition*, 6(3), 315–341.
- Cordingley, E. S. (1989) Knowledge elicitation techniques for knowledge-based systems. In D. Diaper (Ed.), *Knowledge Elicitation: Principles, Techniques and Applications*. John Wiley & Sons, New York.
- Crandall, B., Klein, G. and Hoffman, R. R. (2006) *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. MIT Press, Cambridge, MA.
- d'Aquin, M., Motta, E., Sabou, M., Angeletou, S., Gridinoc, L., Lopez, V. and Guidi, D. (2008) Toward a new generation of semantic web applications. *Intelligent Systems*, 23(3), 20–28.
- Derbentseva, N. and Mandel, D. R. (2011) A concept map knowledge model of intelligence analysis (Ref: TR 2011-077). Defence Research and Development Canada (DRDC), Toronto, Ontario, Canada.
- Dreyfus, H. L. and Dreyfus, S. E. (1986) *Mind over Machine: The Power of Human Intuition and Expertise in an Era of the Computer*. Free Press, New York.
- Duda, J., Gaschnig, J. and Hart, P. (1979) Model design in the PROSPECTOR consultant system for mineral exploration. In D. Michie (Ed.), *Expert Systems in the Micro-Electronic Age*. Edinburgh University Press, Edinburgh, U.K.
- Elstein, A. S., Shulman, L. S. and Sprafka, S. A. (1978) *Medical Problem Solving: An Analysis of Clinical Reasoning* (Vol. 2). Harvard University Press, Cambridge, MA.
- Ericsson, K. A. and Simon, H. A. (1996) *Protocol Analysis: Verbal Reports as Data*. MIT Press, Cambridge, MA.
- Erlandsson, M. and Jansson, A. (2007) Collegial verbalisation – a case study on a new method on information acquisition. *Behaviour & Information Technology*, 26(6), 535–543.
- Erlandsson, M. and Jansson, A. (2013) Verbal reports and domain-specific knowledge: A comparison between collegial and retrospective verbalisation. *Cognition, Technology and Work*, 15(3), 239–254.
- Eskridge, T. C. and Hoffman, R. (2012) Ontology creation as a sensemaking activity. *Intelligent Systems*, 27(5), 58–65.
- Farrington-Darby, T. and Wilson, J. R. (2006) The nature of expertise: A review. *Applied Ergonomics*, 37(1), 17–32.
- Feldman, R. (2013) Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82–89.
- Feltoovich, P. J., Ford, K. M. and Hoffman, R. R. (Eds.) (1997) *Expertise in Context: Human and Machine*. MIT Press, Cambridge, MA.
- Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., Murdock, J. W., Nyberg, E. and Prager, J. (2010) Building Watson: An overview of the DeepQA project. *AI Magazine*, 31(3), 59–79.
- Fransella, F. and Bannister, D. (1977) *A Manual for Repertory Grid Technique*. Academic Press, London, U.K.
- Fransella, F., Bell, R. and Bannister, D. (2003) *A Manual for Repertory Grid Technique* (2nd edn.). John Wiley & Sons, Inc., Chichester, U.K.
- Gaines, B. R. (2013) Knowledge acquisition: Past, present and future. *International Journal of Human-Computer Studies*, 71(2), 135–156.
- Gaines, B. R. and Shaw, M. L. G. (1997) Knowledge acquisition, modelling and inference through the World Wide Web. *International Journal of Human-Computer Studies*, 46(6), 729–759.
- Gammack, J. G. (1987) Different techniques and different aspects of declarative knowledge. In A. L. Kidd (Ed.), *Knowledge Acquisition for Expert Systems: A Practical Handbook*. Plenum Press, New York.
- Gammack, J. G. and Young, R. M. (1985) Psychological techniques for eliciting expert knowledge. In M. Bramer (Ed.), *Research and Development in Expert Systems*. Cambridge University Press, Cambridge, U.K.
- Gavrilova, T. and Andreeva, T. (2012) Knowledge elicitation techniques in a knowledge management context. *Journal of Knowledge Management*, 16(4), 523–537.
- Gil, Y. (2011) Interactive knowledge capture in the new millennium: How the semantic web changed everything. *The Knowledge Engineering Review*, 26(1), 45–51.
- Gray, R. (2004) Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, 10(1), 42–54.
- Grover, M. D. (1983) A pragmatic knowledge acquisition methodology. *Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, Germany.
- Grozniak, V., Guid, M., Sadikov, A., Možina, M., Georgiev, D., Kragelj, V., Ribarič, S., Pirtošek, Z. and Bratko, I. (2013) Elicitation of neurological knowledge with argument-based machine learning. *Artificial Intelligence in Medicine*, 57(2), 133–144.

- Hart, A. (1986) *Knowledge Acquisition for Expert Systems*. Kogan Page, London, U.K.
- Hayes, P., Eskridge, T. C., Saavedra, R., Reichherzer, T., Mehrotra, M. and Bobrovnikoff, D. (2005) Collaborative knowledge capture in ontologies. *Third International Conference on Knowledge Capture*. Banff, Alberta, Canada.
- Hayes-Roth, F., Waterman, D. A. and Lenat, D. B. (1983) *Building Expert Systems*. Addison-Wesley, Reading, MA.
- Heath, T. and Bizer, C. (2011) Linked data: Evolving the web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1–136.
- Hoffman, R. R. (1987) The problem of extracting the knowledge of experts from the perspective of experimental psychology. *AI Magazine*, 8(2), 53–66.
- Hoffman, R. R. (1989) A survey of methods for eliciting the knowledge of experts. *ACM SIGART Bulletin*, 108, 19–27.
- Hoffman, R. R. (1998) How can expertise be defined? Implications of research from cognitive psychology. In R. Williams, W. Faulkner and J. Fleck (Eds.), *Exploring Expertise*. Macmillan, New York.
- Hoffman, R. R., Coffey, J. W., Ford, K. M. and Carnot, M. J. (2001) Storm-LK: A human-centered knowledge model for weather forecasting. *Human Factors and Ergonomics Society Annual Meeting*, Minneapolis, MN.
- Hoffman, R. R., Crandall, B. and Shadbolt, N. R. (1998) Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, 40(2), 254–276.
- Hoffman, R. R. and Lintern, G. (2006) Eliciting and representing the knowledge of experts. In K. A. Ericsson, N. Charness, P. Feltoich and R. R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press, New York.
- Hoffman, R. R., Shadbolt, N. R., Burton, A. M. and Klein, G. (1995) Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes*, 62(2), 129–158.
- Hutchins, E. (1995) *Cognition in the Wild*. MIT Press, Cambridge, MA.
- Jankowicz, D. (2003) *The Easy Guide to Repertory Grids*. John Wiley & Sons Ltd., Chichester, U.K.
- Jankowicz, D. and Thomas, L. (1982) An algorithm for the cluster analysis of repertory grids in human resource development. *Personnel Review*, 11(4), 15–22.
- Johnson, L. and Johnson, N. (1987) Knowledge elicitation involving teachback interviewing. In A. Kidd (Ed.), *Knowledge Elicitation for Expert Systems: A Practical Handbook*. Plenum Press, New York.
- Johnson, P. E., Zualkernan, I. and Garber, S. (1987) Specification of expertise. *International Journal of Man-Machine Studies*, 26(2), 161–181.
- Kelly, G. A. (1955) *The Psychology of Personal Constructs*. W. W. Norton and Company, New York.
- Klein, G. A., Calderwood, R. and Macgregor, D. (1989) Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, 19(3), 462–472.
- Knublauch, H., Ferguson, R. W., Noy, N. F. and Musen, M. A. (2004) The Protégé OWL plugin: An open development environment for semantic web applications. *Third International Semantic Web Conference (ISWC'04)*, Hiroshima, Japan.
- Knublauch, H., Horridge, M., Musen, M., Rector, A., Stevens, R., Drummond, N., Lord, P., Noy, N. F., Seidenberg, J. and Wang, H. (2005) The Protégé OWL experience. *OWL Experiences and Directions Workshop at ISWC2005*, Galway, Ireland.
- LaFrance, M. (1987) The knowledge acquisition grid: A method for training knowledge engineers. *International Journal of Man-Machine Studies*, 26(2), 245–255.
- Maedche, A. and Staab, S. (2003) Ontology learning. In S. Staab and R. Studer (Eds.), *Handbook on Ontologies*. Springer-Verlag, Berlin, Germany.
- Mendonça, F. M., Coelho, K. C., de Andrade, A. Q. and Almeida, M. B. (2012) Knowledge acquisition in the construction of ontologies: A case study in the domain of hematology. *Third International Conference on Biomedical Ontology*, Graz, Austria.
- Milton, N. (2003) *Personal Knowledge Techniques*. University of Nottingham, Nottingham, U.K.
- Milton, N. (2012) Acquiring knowledge from subject matter experts. In J. Kantola and W. Karwowski (Eds.), *Knowledge Service Engineering Handbook*. CRC Press, Boca Raton, FL.
- Milton, N., Shadbolt, N., Cottam, H. and Hammersley, M. (1999) Towards a knowledge technology for knowledge management. *International Journal of Human-Computer Studies*, 51(3), 615–641.
- Milton, N. R. (2007) *Knowledge Acquisition in Practice: A Step-by-Step Guide*. Springer, London, U.K.
- Mitchell, T. M. (1997) *Machine Learning*. McGraw-Hill, New York.
- Mozina, M., Guid, M., Krivec, J., Sadikov, A. and Bratko, I. (2008) Fighting knowledge acquisition bottleneck with argument based machine learning. *Eighteenth European Conference on Artificial Intelligence*, Patras, Greece.



- Newell, A. and Simon, H. A. (1963) GPS, a program that simulates human thought. In E. Feigenbaum and J. Feldman (Eds.), *Computers and Thought*. McGraw-Hill, New York.
- Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, New York.
- Novak, J. and Cañas, A. J. (2006) *The Theory Underlying Concept Maps and How to Construct Them*. Florida Institute for Human and Machine Cognition, Pensacola, FL.
- O'Hara, K., Shadbolt, N. R. and Van Heijst, G. (1998) Generalized directive models: Integrating model development and knowledge acquisition. *International Journal of Human-Computer Studies*, 49(4), 497–522.
- O'Hare, D., Wiggins, M., Williams, A. and Wong, W. (1998) Cognitive task analyses for decision centred design and training. *Ergonomics*, 41(11), 1698–1718.
- Pang, B. and Lee, L. (2008) Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
- Sarawagi, S. (2008) Information extraction. *Foundations and Trends in Databases*, 1(3), 261–377.
- Schreiber, G. (2013) Knowledge acquisition and the web. *International Journal of Human-Computer Studies*, 71(2), 206–210.
- Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N. R., Van de Velde, W. and Weilinga, B. (2000) *Knowledge Engineering and Management: The CommonKADS Methodology*. MIT Press, Cambridge, MA.
- Schvaneveldt, R., Durso, F., Goldsmith, T., Breen, T., Cooke, N., Tucker, R. and Maio, J. (1985) Measuring the structure of expertise. *International Journal of Man-Machine Studies*, 23(6), 699–728.
- Schweikert, R., Burton, A. M., Taylor, N. K., Corlett, E. N., Shadbolt, N. R. and Hedgecock, A. P. (1987) Comparing knowledge elicitation techniques: A case study. *Artificial Intelligence Review*, 1, 245–253.
- Scott, C. A., Clayton, J. E. and Gibson, E. L. (1991) *A Practical Guide to Knowledge Acquisition*. Addison Wesley, Boston, MA.
- Shadbolt, N. and Burton, A. M. (1989) The empirical study of knowledge elicitation techniques. *ACM SIGART Bulletin – Special Issue on Knowledge Acquisition*, 108, 15–18.
- Shadbolt, N., Hall, W. and Berners-Lee, T. (2006) The semantic web revisited. *IEEE Intelligent Systems*, 21(3), 96–101.
- Shadbolt, N. and Milton, N. (1999) From knowledge engineering to knowledge management. *British Journal of Management*, 10(4), 309–322.
- Shadbolt, N. R. (2005) Eliciting expertise. In J. R. Wilson and N. Corlett (Eds.), *Evaluation of Human Work* (3rd edn.). CRC Press, Boca Raton, FL.
- Shadbolt, N. R. (2013) Knowledge acquisition and the rise of social machines. *International Journal of Human-Computer Studies*, 71(2), 200–205.
- Shadbolt, N. R. and Burton, A. M. (1990a) Knowledge elicitation. In J. R. Wilson and E. N. Corlett (Eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology* (2nd edn., pp. 406–440). Taylor & Francis Group, London, U.K.
- Shadbolt, N. R. and Burton, A. M. (1990b) Knowledge elicitation techniques: Some experimental results. In K. L. McGraw and C. R. Westphal (Eds.), *Readings in Knowledge Acquisition*. Ellis Horwood, New York.
- Shadbolt, N. R. and Burton, M. (1995) Knowledge elicitation. In J. R. Wilson and N. Corlett (Eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology*. Taylor & Francis Group, London, U.K.
- Shadbolt, N. R., O'Hara, K. and Crow, L. (1999) The experimental evaluation of knowledge acquisition techniques and methods: History, problems and new directions. *International Journal of Human-Computer Studies*, 51(4), 729–755.
- Shaw, M. L. G. (1980) *On Becoming a Personal Scientist: Interactive Computer Elicitation of Personal Models of the World*. Academic Press, London, U.K.
- Shaw, M. L. G. and Gaines, B. R. (1987) An interactive knowledge elicitation technique using personal construct technology. In A. L. Kidd (Ed.), *Knowledge Acquisition for Expert Systems: A Practical Handbook*. Plenum Press, New York.
- Shaw, M. L. G. and Gaines, B. R. (2001) WebGrid: Knowledge elicitation and modeling on the world wide web. In R. Rajkumar (Ed.), *Industrial Knowledge Management: A Micro-Level Approach* (pp. 335–348). Springer-Verlag, London, U.K.
- Shortliffe, E. H. (1976) *Computer-Based Medical Consultations: MYCIN*. Elsevier, New York.
- Stewart, T. A. (1997) *Intellectual Capital: The New Wealth of Organizations*. Nicholas Brealey, London, U.K.
- Sure, Y., Staab, S. and Studer, R. (2003) On-To-Knowledge Methodology (OTKM). In S. Staab and R. Studer (Eds.), *Handbook on Ontologies*. Springer-Verlag, Berlin, Germany.
- Tudorache, T., Nyulas, C., Noy, N. F. and Musen, M. A. (2013) WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web. *Semantic Web*, 4(1), 89–99.

- Van Gog, T., Paas, F. and Van Merriënboer, J. J. G. (2005) Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, 19(2), 205–221.
- Walker, B. M. and Crittenden, N. (2012) The use of laddering: Techniques, applications and problems. In P. Caputi, L. L. Viney, B. M. Walker and N. Crittenden (Eds.), *Personal Construct Methodology*. Wiley-Blackwell, Chichester, U.K.
- Wang, Y., Sure, Y., Stevens, R. and Rector, A. (2006) Knowledge elicitation plug-in for Protégé: Card sorting and laddering. *First Asian Semantic Web Conference*, Beijing, People's Republic of China.
- Weiss, S. and Kulikowski, C. (1984) *A Practical Guide to Designing Expert Systems*. Rowman & Allanheld, Totowa, NJ.
- Witten, I. H. and Eibe, F. (2005) *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, San Francisco, CA.
- Wood, L. E. and Ford, J. M. (1993) Structuring interviews with experts during knowledge elicitation. In K. M. Ford and J. M. Bradshaw (Eds.), *Knowledge Acquisition as Modeling*. Wiley, New York.
- Wu, H., Zubair, M. and Maly, K. (2006) Harvesting social knowledge from folksonomies. *Seventeenth Conference on Hypertext and Hypermedia*, Odense, Denmark.
- Zsombok, C. E. and Klein, G. A. (Eds.) (1997) *Naturalistic Decision Making*. Lawrence Erlbaum Associates, Mahwah, NJ.



---

# 8 Simulation and Digital Human Modelling

*Glyn Lawson and Gary Burnett*

## CONTENTS

|                                                                                   |     |
|-----------------------------------------------------------------------------------|-----|
| Introduction.....                                                                 | 201 |
| Simulation and DHM within the Human–Machine System .....                          | 202 |
| Primary Application Areas and Industries .....                                    | 202 |
| Fidelity and Validity in Human Factors Simulators.....                            | 206 |
| Sickness in Simulators .....                                                      | 208 |
| Constructive Simulation: DHM Software Packages and Modelled Characteristics ..... | 209 |
| Anthropometry Modelling .....                                                     | 210 |
| Biomechanical Modelling.....                                                      | 211 |
| Anatomical Modelling .....                                                        | 211 |
| Modelling the Human Mind.....                                                     | 212 |
| Evaluation Tools Provided within DHM.....                                         | 212 |
| Advantages and Disadvantages of Simulation and DHM.....                           | 212 |
| The Future.....                                                                   | 214 |
| Guidelines for Using Simulation and DHM .....                                     | 214 |
| Conclusions.....                                                                  | 215 |
| References.....                                                                   | 216 |

## INTRODUCTION

Simulation as a concept has been around for most of the history of humanity, for example as a means of planning activities such as battles before deploying resources (strategic use), for providing personnel with appropriate skills in a safe/controlled environment (training use) and to allow people to escape reality to engage in fictional worlds (entertainment use) (Vincenzi et al., 2009). In the modern era, computer-based simulation is now widely used in ergonomic and human factors (E/HF) research and practice as a cost-effective means of investigating peoples' interactions with technology (Wickens et al., 2004). It can support predictions of the consequences of a design, allow detailed investigation of human performance or support training for scenarios which are too dangerous in the real world. Simulation purports to replicate elements from the real world and also has an element of pretence. Both aspects raise fundamental questions regarding both the characteristics of a simulation for the purposes of E/HF study (how real does it have to be, and what elements are actually important?) and the subsequent performance and behaviour observed within a simulation environment (how do people respond to the pretence of a simulation?). This chapter addresses these questions by considering the broad concept of simulation, together with the more specific area of Digital Human Modelling (DHM) in which digital representations of humans, developed using human data to be as realistic as possible, are used within a virtual environment for ergonomics analysis (Demirel and Duffy, 2007; Landau, 2006). We provide an overview of current uses and issues with simulation and DHM, including a discussion of validity and fidelity. Information is

provided about the various DHM packages available for purchase and the functionalities offered by these tools. After a summary of the pros and cons of simulation-based approaches, suggestions are made for the future of simulation and DHM. The chapter ends with guidelines for using these tools for E/HF investigations.

## SIMULATION AND DHM WITHIN THE HUMAN–MACHINE SYSTEM

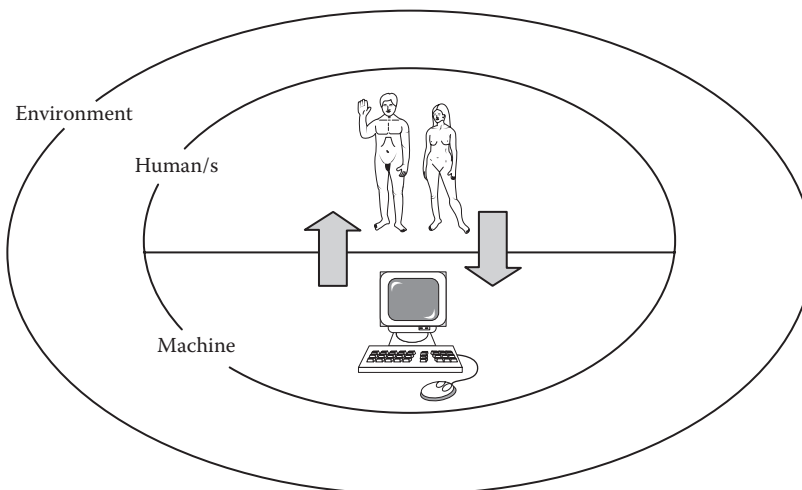
The human–machine system (see Figure 8.1) is a useful means for us to consider what characteristics of a system may be represented in digital form. Essentially, we may choose to simulate combinations of human, machine or environment (physical, social or organisational). The particular components of the system that are simulated have been referred to as the dimensions of simulation (Andrews et al., 2010) and include the following:

- ‘Live’ simulation – real people using a simulated machine in the real environment, for example a user trial with a prototype interface.
- ‘Virtual’ simulation – real people using a simulated machine in a simulated environment. There may be some physical elements of the environment, for example flight and car simulators or the environment may be completely virtual, for example CAVE.
- ‘Constructive’ simulation – simulated people using a simulated machine in a simulated environment, for example DHM software, or autonomous agents (computer-generated people with their own behaviours not under direct control of the operator) in a virtual building.

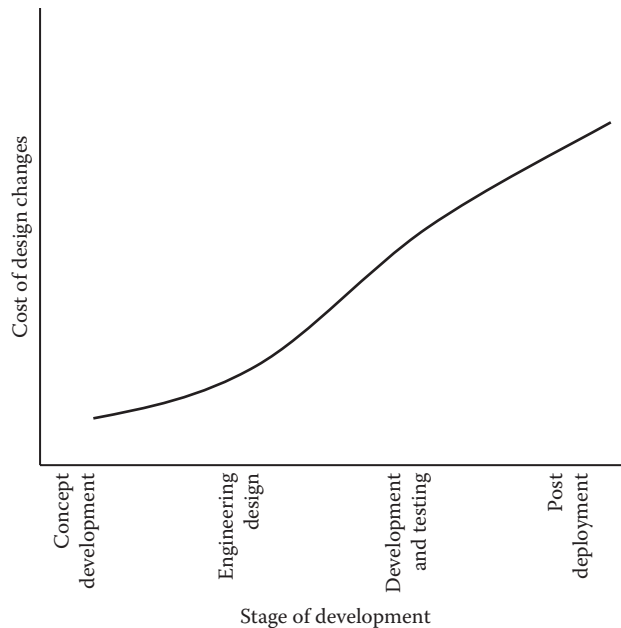
More detail can be added to the Andrews et al. classification to include other pertinent issues. For instance, a simulation will be defined largely by its purpose as well as the ‘users’ of the simulation. Users may not only include operators in the conventional sense, but also incorporate designers, managers and those with a marketing perspective. As we will see later in this chapter, different types of user will have very different needs associated with the simulation.

## PRIMARY APPLICATION AREAS AND INDUSTRIES

There are many application areas and industries that utilise simulation and DHM approaches and several arguments for their use such as cost savings, standardising evaluation protocols and minimising danger to users. In design and engineering, they have been used in the automotive industry



**FIGURE 8.1** The human–machine system.



**FIGURE 8.2** Cost of design changes by development stage. (From Laughery, R., *Simulation and modelling for analysing human performance*, in Wilson, J.R. and Corlett, N. (eds.), *Evaluation of Human Work*, 3rd edn., Taylor & Francis Group, London, U.K., 2005, pp. 219–238.)

(e.g. Bowman, 2001; Chaffin, 2005; Sabbah et al., 2009; Zehner and Robinette, 2007), for workplace evaluation and design (e.g. Albers and Abshire, 1998; Chaffin, 2005; Chang and Wang, 2007; Du and Duffy, 2007; Hanson et al., 2012; Lamkull et al., 2009), for aircraft design (e.g. Karmakar et al., 2012; Punte et al., 2001) and for other sectors. As mentioned, simulation and DHM can lead to a reduction in cost in the design and engineering process. This is in part due to the resource savings associated with a proactive approach to E/HF: by identifying and resolving ergonomics issues early in the process, costly rework later in the design/engineering process can be reduced or avoided (Bowman, 2001; Chaffin, 2001; Duffy, 2009; Laughery, 2005). This principle is nicely illustrated by the graph in Figure 8.2 and by example; it is easy to see how modifying a car with insufficient headroom would cost little at the concept stage following a DHM evaluation; the costs of this change this would be much higher if this issue was identified only after the car rolls off the production line. Of course, it is unlikely that a car would be produced without anyone driving a prototype between concept stage and deployment, which itself raises another important point that DHM can support design by predicting human interaction with a system but should not replace testing with real humans using physical prototypes. There are also cost savings associated with the use of simulation and DHM for ergonomics analysis through a reduction in the number of physical prototypes which must be built and tested with human participants (Duffy, 2009; Laughery, 2005). Testing with users can be expensive (Neilson, 1993); commercial organisations will welcome opportunities to reduce the number of times this cost is incurred. The reduced time per design iteration through the use of simulation and DHM can also increase the number of design solutions investigated (Laughery, 2005) which may ultimately result in a more satisfactory solution for the end-users.

DHM may also address the problems of recruiting and testing the full range of target users. Several manikins with multiple anthropometric configurations (i.e. short body/long legs, long body/short legs, etc.) can be evaluated relatively quickly and easily (Duffy, 2009) – more so than with human participants, if in fact such a test panel of real users could be found and recruited. Using simulated people

can also avoid issues with recruiting a sample of people to evaluate a design which may be commercially sensitive or protected by security clearance. DHM can also address difficulties in applying scientific literature and recommendations/guides to design, as they can be too general or hard to interpret. For example, it is often the case when designing a product that the standard postures which are used for anthropometric measurement (e.g. standing erect looking ahead with arms at sides, or sitting with thighs horizontal and feet flat on the floor) do not apply – DHM can combine anthropometric and biomechanical data to create the posture of interest. DHM can also be used as part of a standardised evaluation methodology (Hanson et al., 2006). There may be a desire to use DHM models as visual representations of humans to see how the user or customer may look when interacting with a product (Duffy, 2009) – this can not only be used for E/HF analysis (e.g. to check the practicability of a proposed maintenance manoeuvre) but also be used in marketing or supporting discussions within the organisation. The use of a visual image showing a DHM manikin interacting with a proposed design can sometimes be very powerful in making an argument for a change. Finally, DHM offers support for many E/HF analysis methods, for example, by analysing the postures of the manikin and providing additional details about the task, the DHM software may be able to calculate the level of musculoskeletal risk according to RULA or the recommended weight limit based on the NIOSH equation (see Chapters 15 through 17) which may be easier than with a paper-based exercise.

Simulation may benefit design applications in situations which are too dangerous to recreate in real life. For example, driving simulators are often a preferred means of assessing a new user interface, for which the distraction effects on the driver are unclear (e.g. Burnett et al., 2013) and therefore a road-based assessment would be unethical. Simulations have also been used to determine the safety of buildings during emergency evacuation, as they can incorporate smoke and fire, which would be too dangerous to create in real life (Ren et al., 2008). This latter scenario could be implemented either as virtual simulation with a person controlling an avatar or as a constructive simulation populated entirely by autonomous agents. The advantages include being able to investigate safety early in the design phase, and if a constructive simulation (possibly populated by autonomous agents with some form of artificial intelligence) is used, it can be run many times with minimal effort to understand the potential distribution of evacuation times, rather than one instance obtained from a fire drill (Gwynne et al., 1999).

Training is another established use for simulation. The first author of this chapter worked on the development of an emergency training simulator for the DiFac Digital Factory for Human-Oriented Production EU-funded research project (FP6-IST-5-035079) (Figure 8.3). The aim was to train people on a desktop virtual simulator for emergency situations. Thus, the simulator enabled training for a dangerous situation without any risk of physical harm. Another motivation for the development of the simulator was that fire drills (a common approach to training for real evacuations) result in people being away from their jobs and therefore reduce productivity – this simulator meant that people could train individually and at their convenience without stopping the production line. Further advantages to the simulator were that the training session could be reviewed afterwards from any viewpoint to give the trainee an appreciation of the contexts and broader consequences of their actions. The training session could also be fast-forwarded to critical points and reviewed as many times as necessary.

Another virtual training system was developed on the VISTRA Virtual Simulation and Training of Assembly and Service Processes EU funded project (ICT-285176) [www.vistra-project.eu](http://www.vistra-project.eu) (Figure 8.4). The VISTRA project aimed to develop a training simulator to allow assembly line operators to train on virtual cars before prototypes of vehicles become available. It uses a body motion tracking system to capture the trainees' postures, which is linked to a virtual representation of the tool and car assembly. The advantages to this approach include

- Ability to train on a greater number of vehicle variants as typically only a small number of physical prototypes are used during automotive assembly training.
- A potential reduction in the costs of training on hardware, including travel time to dedicated training centres plus the costs of degradation of the physical prototypes through



**FIGURE 8.3** Image of an emergency training simulator developed on the DiFac EU-funded project (FP6-IST-5-035079), based on SWORD Critical Infrastructure. (Image courtesy of MASA Group S.A.)



**FIGURE 8.4** VISTRA training solution.

repeated assembly and disassembly. Disassembly of the prototypes is actually wasteful (time and cost) as the main aim of the activity is training for assembly.

- Ability to train earlier in the vehicle development cycle (rather than waiting for physical prototypes to become available) thus reducing the time to launch a vehicle.
- Use of relatively inexpensive technologies which can feasibly be deployed trackside rather than a dedicated training facility which would require trainees to move away from their workplaces. The hardware used takes advantages of the recent affordability of body tracking technologies (e.g. Microsoft Kinect® for Windows).

Research conducted for this project found virtual training resulted in reductions in total training time compared to traditional approaches, as well as more positive attitudes towards the virtual training experience (Baines, 2012). Further testing at the industrial user partners' sites demonstrated a reduction in task error following training on the virtual system compared to traditional paper-based training. However, operators commented that the virtual system should supplement rather than replace training on hardware to ensure that aspects such as 'the feel of the parts' are still learnt.

Other applications for simulation-based training include driver training (e.g. Wang et al., 2010), medical training (e.g. Stone, 2011) and training airport crews in handling airline crashes (Weiss and Jessel, 1998), amongst others. In each of these, there is a safety issue associated with the situation for which the trainee is being trained – simulation can reduce exposure to that danger.

Simulators also offer advantages in research including high control over variables which would be difficult or impossible to control in the real-world environment. For example, in a train simulator, each driver can experience the same section of track in exactly the same environmental conditions, a consistency that could not be guaranteed in real-world trials. This allows performance to be monitored in response to the events of interest. A related point is that the measurement of various dependent variables could be difficult in the real world, for example eye tracking may prove unreliable, physiological monitoring may prove intrusive, or interrupting the train driver to ask for subjective ratings may prove dangerous.

## FIDELITY AND VALIDITY IN HUMAN FACTORS SIMULATORS

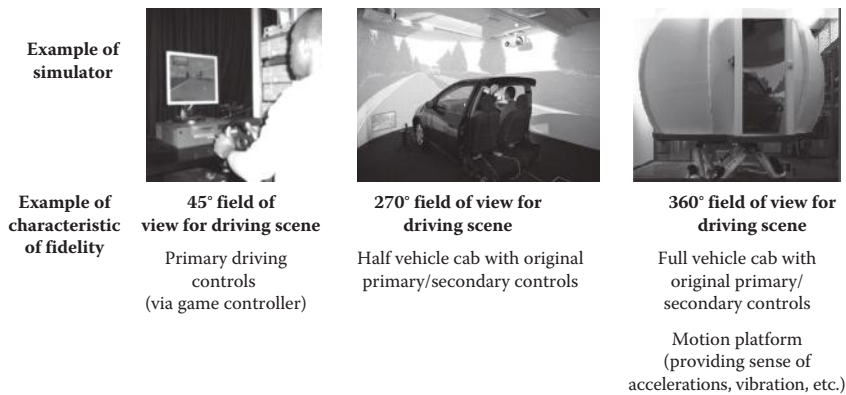
Vincenzi et al. (2009, p. 21) define simulation fidelity as 'the degree to which a model or simulation reproduces the state and behaviour of a real world object or perception of a real world object, feature, condition or standard in a measurable or perceivable manner'. A key inclusion in this definition is the word 'perception', recognising the fact that the operator's subjective impression of a simulation will help to define its success, rather than purely the objective measurement of how close the technology can reproduce reality. Fidelity is a critical topic for simulation for two key reasons. First, fidelity is closely linked to cost as, in most cases, higher fidelity requires more sophisticated equipment (e.g. higher-resolution projectors, surround sound systems, motion platforms). Second, it is well established that the fidelity of a simulation will have a considerable impact on an operator's motivation, performance and behaviour (e.g. Burnett et al., 2007). As a consequence, it is clear that the ultimate cost–benefit ratio for a simulator can be defined largely by its fidelity.

In this consideration, it is important to note that fidelity is *not* the same as validity. While higher fidelity is often associated with more realistic performance/behaviour for participants, the validity of a simulation study is also affected by many other things, for example procedures, choice of participants, study design and what/how you measure. From an E/HF perspective, an overall recommendation in designing a simulation experience is to reflect on the fundamental purpose of the simulation and understand what aspects of fidelity are critical to task behaviour/performance. For instance, in a driving simulator study of the distraction caused by a new navigation system interface, there would be little need to show intricate details of a front door of a house that drivers were passing. In contrast, it would be much more important to provide a realistic, high-fidelity representation of surrounding traffic or particular road signs.

Several authors have distinguished between types of fidelity for human factors simulation, primarily considering the distinction between physical and psychological components (e.g. Reed and Green, 1999). In this respect, physical fidelity refers to whether a simulation looks/sounds/feels like the real world and thus relates to the accuracy and completeness with which a simulation provides input to the human senses. Examples of three driving simulators that increase (in general terms) in fidelity from left to right are shown in Figure 8.5.

As a specific form of physical fidelity, audio–visual fidelity is arguably the most developed in modern simulators and may provide geo-specific (e.g. a particular landscape or object that exists





**FIGURE 8.5** Physical fidelity differences in driving simulators.

in the real world) or geo-typical (a generic world) representations. Geo-specific fidelity will be required in certain situations, for example training pilots for approaches to a known runway or military operations in a particular area – but will need to be very accurate, as the consequences of differences between real and simulated worlds may be considerable (e.g. Bach et al., 2008). In contrast, a geo-typical world requires an understanding of the basic characteristics needed for appropriate task performance, for example a certain number of lanes of given width on motorway, and thus can be of lower overall fidelity.

Motion fidelity concerns the extent to which a simulation reproduces the sense of motion felt in the real world and is notoriously difficult to represent accurately and convincingly in a simulation. This is partly because the body utilises many sensory receptors for ‘true’ motion perception, including proprioceptors (such as in muscles, tendons or joints), the semicircular canals (in inner ear) and vestibular sacs (also in inner ear). A simulation may use mechanisms, usually within hexapod systems, to ‘trick’ the body into experiencing a sense of motion, for example by quickly tilting a flight simulator forward to give a sense of deceleration. It is commonly the case that considerable efforts are required in the calibration and fine-tuning of these systems to avoid negative effects, such as symptoms of sickness.

There have been few studies considering the validity of simulations that include motion systems. These include comparisons of driver behaviour in motion-based simulators to behaviour in the real world (e.g. Godley et al., 2002; Yan et al., 2008). These studies have demonstrated relative validity, that is behaviours in the same order effect as the real world, but lack absolute validity. It is commonly the case that simulation should be considered as having greater relative validity, that is for comparing designs, than absolute validity, that is transferring the results to the real world (Reuding and Meil, 2004; Naghiyev et al., 2014), although this depends on the specifics of the task, equipment and people of interest. Interestingly, in training for pilots (especially novice pilots in standard airplanes), there has been some research showing few benefits for the use of sophisticated motion systems (Vincenzi et al., 2009). Nevertheless, for more complex situations (e.g. fighter pilots training for complex manoeuvres), motion has been demonstrated to assist in training. Vincenzi et al. (2009) hypothesise that such advantages arise because expert users require a wider range of cues for successful completion of a task and also have greater expectations for a task. In contrast, there is a view that too much fidelity can distract from training goals for novices. Unfortunately, the lack of understanding across disciplines and simulator uses makes it difficult to generalise from these conclusions.

A second type of fidelity with implications for validity concerns the psychological aspects of a simulation and considers whether the simulator replicates the relevant cognitive factors involved. Here, the focus is on the operator within a simulation (rather than the machine/s) and can concern



individual and task variables. For the former, it is important to consider an operator's perception of a simulator and where they believe they currently are in physical space, together with the activities being conducted. In the virtual reality literature, this concept is commonly referred to as 'presence' – defined by Witmer and Singer (1998, p. 225) as 'the subjective experience of being in one place or environment, even when one is physically situated in another'.

When considering task variables relevant to the fidelity of a simulator, there is a requirement to understand *what* an operator might need to do (i.e. which tasks are of interest). For instance, when training nurses in an operating theatre context, fidelity would be higher when all the relevant people are present (either physically or via a communication link). There is also a need to consider *how* a task should be carried out (i.e. can an operator undertake the task as they would in the real world?). As an example, a driving simulator for use in the United Kingdom would require the option for manual gear changes, given the prevalence of this mode of operation for the population.

Returning to the concept of validity, greater levels of fidelity may contribute to higher face validity, which is the appearance of a method being real, at face value (Banyard and Grayson, 2000) (and see Chapter 1 for further discussion of validity). While face validity has been criticised as a weak measure of validity due to the subjective value of the experimenter, it may provide an indication of whether the simulator is measuring what it is supposed to (Howitt and Cramer, 2011). Moreover, face validity from the stakeholder's perspective is important – if a senior manager does not believe the simulator is a faithful replication, they will be unwilling to accept the results obtained from it.

## SICKNESS IN SIMULATORS

A well-known side effect of immersion within virtual reality-based simulators is sickness, associated with a range of symptoms, such as headache, eyestrain, dizziness, fatigue and nausea. Other negative effects of simulator use include changes in oculomotor response and postural instability (Nichols and Patel, 2002). Clearly, the well-being of those people within a simulator is of importance to human factors researchers and practitioners. In addition, the utility of a simulator can be reduced significantly if certain individuals are not able to make use of the facility which is critical in the area of training/assessment. Finally, there is a close link between sickness and validity, as people experiencing sickness in a simulator have been shown to negatively modify their behaviour as a coping mechanism (e.g. closing eyes when cornering in a driving simulator, Burnett et al., 2007; Nichols et al., 2000). Risk factors for sickness in human factor simulators can be grouped into those relevant to simulator design (fidelity); the type of study (tasks and exposure) and individual differences (susceptibility). With respect to the former, a fundamental factor concerns whether a simulator can lead to conflict, for example between what the visual and vestibular senses are informing the brain on user movement within a virtual world (Reason and Brand, 1975). Furthermore, certain tasks create greater risk for sickness (e.g. those where optic flow is higher, such as turning at a junction in a driving simulator). As a third issue, it is clear that some people are more susceptible to sickness than others (e.g. Park et al., 2006). However, due to conflicting results in the literature, there appears to be no consensus as to whether certain groups (e.g. older people) are at greater risk than others (Sharples et al., 2014).

Intervention strategies for preventing simulator sickness tend to focus on a combination of approaches, largely aiming to minimise the likelihood of symptoms developing. These strategies should attempt to address the technology (e.g. high projection quality, avoid sensory conflict); the environment (e.g. aim to minimisevection – feelings of self-motion and avoid large amounts of movement); the task (e.g. habituation through repeated exposure can minimise sickness) and the individuals through selection (e.g. avoid recruiting participants with susceptibility to motion sickness) (Nichols and Patel, 2002). In addition, researchers and practitioners should employ good practice to reduce the consequences of sickness should it arise, such as having a quiet, cool place away from the simulator to relax, providing drinking water and affording adequate recovery time before driving or operating heavy machinery.

## CONSTRUCTIVE SIMULATION: DHM SOFTWARE PACKAGES AND MODELLED CHARACTERISTICS

The previous sections focused on ‘virtual’ simulation; however, fidelity (in terms of realism) and validity concerns also affect DHM and other forms of constructive simulation. These issues will be discussed after an introduction to the available DHM packages. Thereafter, the functionalities offered within the many DHM tools are described.

There are over 100 DHM software packages available. These vary according to the accuracy and completeness of the models, specificity of application area and cost. In general, the state-of-the-art DHM tools are more advanced in modelling physical aspects of the human rather than behavioural/cognitive aspects, most likely due to the complexity of modelling the human mind. Thus, DHM is often primarily employed to evaluate workplaces or *hardware*, rather than *software* or information technology (Alexander and Gunther, 2014).

Some of the generic and widely used packages include JACK, SAMMIE, SANTOS, SAFEWORK and HumanCAD. JACK is a well-developed package offered by Siemens, which has a range of anthropometric and ergonomics analysis features. SAMMIE was developed at Loughborough University and the University of Nottingham and consequently has a proven background within the research community, although is now offered for sale and is the basis of the SAMMIE CAD Ltd. consultancy. SANTOS was originally developed for defence applications at the University of Iowa but is now used in applications including aerospace, transportation and agriculture, and offers many advanced modelling features such as muscle fatigue, physiological systems and clothing modelling. SAFEWORK was developed in the United States in the 1980s based on US army data and is popular within several universities in the United States. As indicated by the title, it has an emphasis on health and safety applications but still offers generic functionality for supporting ergonomics design in a variety of applications. HumanCAD (the successor to ManneQuin) was also generated in the United States and is a popular model offering a wide range of functionalities.

Examples of the more specific packages include RAMSIS and BHMS. RAMSIS is used in vehicle design and offers a posture model based on real humans’ chosen seating positions. Used by around 75% of all vehicle manufacturers, it also offers other functionalities useful for vehicle design such as reach analysis, postural comfort analysis, visibility analysis and seat belt route calculation. BHMS (Boeing Human Modeling System) was developed to support aircraft design and maintenance. It offers the generation of clearance zones required for access and hand-tool modelling for use during maintenance task investigations. It also supports cockpit design with features such as vision assessment and reach zones.

A wide range of characteristics of people can be modelled within the various DHM tools. Some of the more common characteristics include static and functional anthropometry, weight, strength and movement characteristics. However, more specialised programmes can model internal organs/structures, sensory (e.g. vision, hearing, touch) characteristics and individual abilities/limitations in perception, attention, memory and decision making. There are also clothing models, to investigate the additional space requirements, for example for protective clothing, or for the development and sales of clothing within the textiles industry.

Several of the tools include representations of skin, hair and clothing of the modelled humans. While the clothing can be useful for studying additional space requirements, these aspects are generally motivated by the desire to increase the realism of the appearance of the model. However, this trend towards increasing the realism of the appearance and behaviour of the models is not always beneficial. Positive aspects include increasing the face validity and improving the ease of communication of ideas and issues, sometimes making it easier to convince senior management of an ergonomics issue. However, increasing the realistic appearance of manikins may make management too ready to accept a DHM investigation as ‘real’ without understanding the limitations of this approach (e.g. a DHM manikin will not fidget, shuffle or move as a consequence of a prolonged static posture as a real human would; a workplace which fits snugly around a manikin may appear okay to a senior

manager but may feel very constrained to the worker). Moreover, the additional complexity may not add anything to the issues under investigation, instead making the software more complex or at least more computationally demanding and therefore slow. For example, investigations of reach, view to displays and clearance do not depend upon photorealism of the model.

The following section covers the main features and functionality of DHM tools, based on the categorisations provided by Bubb and Fritzsche (2009). It should be noted that in the commercially available software, the models are being developed with increasing functionality and therefore may offer a combination of these features.

## ANTHROPOMETRY MODELLING

Anthropometric models are used to optimise reach, clearance and posture in a variety of applications. These are often where space saving is a priority, for example cockpit and cabin design in the automotive and aviation industries, in which the design is built around the minimum requirement of the largest target human (Bubb and Fritzsche, 2009). Anthropometric models are generally made up of a skeleton model, which uses joints designed to mimic a human skeleton, and a 'skin' or outer layer which is to increase realism and represent the volume required for the human (Bubb and Fritzsche, 2009).

Most anthropometric models are based on anthropometric data tables which are adapted to work within the DHM. They often provide some ways to generate body segment parts based on an overall dimension. That is, the 95th percentile stature person will not necessarily have 95th leg length; the anthropometric model must determine leg length based on the overall stature (Bubb and Fritzsche, 2009). Often, the DHM tools allow the creation of standard manikins (tall, average, short, often based on stature and corresponding to, e.g. 95th, 50th and 5th percentiles) and 'limiting' or 'boundary' manikins for evaluation – for example tall male/long body, tall male/short body, short female/long body and so on (Brolin et al., 2012; Summerskill and Marshall, 2011). In this way, the ergonomics investigator can still investigate several anthropometric dimensions at once, which would be a tedious task without the DHM and relying solely on anthropometric data tables, but be confident that they are not falling foul of the fallacy of the 5th percentile female/95th percentile male, that no such person exists with 5th or 95th percentile measurements for each body part.

A number of anthropometric data sources are used within the DHM tools (Beeney and Charland, 2012; Godil and Ressler, 2009). These include, for example, the ANSUR (Gordon et al., 1989) survey of over 75,000 US military personnel, which was updated in 2009 and is being implemented in the latest versions of DHM software. For civilian data, large-scale anthropometric surveys include National Health and Nutrition Examination Survey of US adults and children conducted by the Centers for Disease Control and Prevention (McDowell et al., 2009). The CAESAR (2006) survey included 5000 participants and combined 3D scanning technology with traditional body measurements. Other recent surveys which have used 3D scanning technologies include Size UK (2006) and Size USA (2006), with over 10,000 participants each. Despite these large-scale surveys, there may be a need to customise the dimensions of any particular manikin (e.g. to match a specific user), and many of the DHM tools offer this functionality. It is important to question the anthropometric data upon which any particular DHM is based; it is all too easy to create a 'default male' or '5th percentile female' without giving consideration to match to your target population (including any scaling for secular growth).

Another important aspect of anthropometric models is *inverse kinematics* – the ability of the models to define a realistic posture based on certain constraints. For example, if a digital human is made to sit at a workstation, and then their hand is attached to some tool, the resulting arm posture is calculated by inverse kinematics (i.e. without specifying each individual body segment). However, the calculated posture is only one of many postures the human may take. Also, the overall position of the manikin is still influenced by the subjectivity of the assessor (De Magestris et al., 2013).

Other concerns include the ‘robotic’ and unrealistic appearance of the motions and forms predicted using inverse kinematics (Chaffin, 2001).

## BIOMECHANICAL MODELLING

Biomechanical modelling involves modelling the dynamic behaviour and function of the human body. The body segments in models offering this functionality may have physical properties such as mass and realistic joint behaviour to represent the moments and forces involved in human movement and exertions (Bubb and Fritzsche, 2009; Chaffin, 2001). The models are populated with data from measurements on real humans and are implemented using mathematical and mechanical force predictions based on laws of physics (Bubb and Fritzsche, 2009; Chaffin, 2001).

Another approach to populating movement data is to track a human in the real world, then digitise their movement paths. This approach is becoming more feasible as body motion tracking reduces in costs. It can avoid time-consuming manual positioning of the manikin for complex tasks such as vehicle entry and exit. However, the captured data often need fairly extensive manipulation to produce realistic motion and may also present problems such as missing data points which were not picked up by the motion capture system (Bradwell and Li, 2008). Moreover, to track the movement of the real human, some physical representation of the workspace of interest may be required – in this case, it is worth questioning whether the evaluation would be better conducted using real humans. DHM may still offer some advantages, such as support for ergonomics assessment methodologies.

The modelling of dynamic behaviour of the human body is important as it can relate to key design and ergonomics criteria, for example maximising comfort or minimising fatigue. The aim is often to understand the impact of an activity on the human body, particularly to evaluate the risk of injury (Woldstad, 2006). Current DHM tools have been criticised for not being able to sufficiently evaluate the forces exerted on humans by their work over time, thus leading to inaccurate assessment of biomechanical risk (De Magestris et al., 2013; Ma et al., 2009). De Magestris et al. (2013) developed a dynamic DHM controller which requires as input only the start point, end point, a mid-point in the movement (to calculate object avoidance) and the force required for the task. It provides as output a dynamic simulation including joint angles and forces, which are reported to correspond to experimental data (De Magestris et al., 2013). Ma et al. (2009) modelled fatigue effects over time, which are currently inadequately modelled in most DHM tools. They also included a joint discomfort parameter, which is based on the range of motion of the joint. In this way, discomfort and fatigue can be investigated as a consequence of workstation layout. Other examples of ergonomics uses of biomechanical models include manual handling investigations; investigating vehicle crash worthiness (to allow, e.g. head impact analysis or whiplash); analysing whole-body vibration; medical investigations, for example into joint behaviour and strains; and ingress egress analysis (Bubb and Fritzsche, 2009; Woldstad, 2006) (see Chapter 17 for further discussion of biomechanical bases of ergonomics assessment).

## ANATOMICAL MODELLING

Anatomical models are computer-generated representations of the human body which depict the internal organs’ volume and densities (Bubb and Fritzsche, 2009; Caon, 2004). These have advanced with the use of medical scanning and imaging equipment which are able to capture digital sections of the internal structure of the human body (Caon, 2004). These data are used to create ‘voxel’ models – a term derived from volumetric and pixel – which represent the 2D sections of organs as 3D models.

Anatomical models have obvious use in medical training or planning for surgery. They are also used to investigate the impact of radiation on organs, for example during radiotherapy (Caon, 2004). A growing body of work aims to model head injury following a car crash, to improve automotive safety or treatment following an incident (e.g. Horgan and Gilchrist, 2004).

## MODELLING THE HUMAN MIND

In addition to modelling the physical elements of a job or workstation, it may also be important to model cognitive aspects. This could, for example, help determine how a task will be done, how long a task may last or identify areas where errors may arise when the demands of the work exceed the capabilities of the human. Bubb and Fritzsche (2009) describe the main categories of models of cognition as *cause–effect models* and *probability models*. The former predicts reductions in performance as a consequence of events occurring in the human–machine interaction. *Probability models* focus on the likelihood of successful outcomes, based on calculation of error probabilities. These are based on theoretical models such as those provided by Reason (1990) and Rasmussen (1987). Other tools such as ACT-R (Anderson et al., 2004) aim to allow for the realistic recreation of aspects of cognition such as memory, learning and decision making.

While it is currently rare that such cognition and psychological aspects are modelled in commercially available DHM systems, it is likely that in the future, elements of the human mind will be increasingly included as more models emerge and as computer processing power continues to improve. Their usefulness in DHM is obvious, for example determining which strategy a human may use to approach a task, error likelihood, emotional reactions and the impact of other behaviours (e.g. procrastination) on task performance. There is a strong argument for considering both physical and psychological aspects together given the interaction between them for many ergonomics considerations, for example motivation to do a job can impact on the likelihood of musculoskeletal disorders.

## EVALUATION TOOLS PROVIDED WITHIN DHM

DHM tools commonly offer a number of evaluation tools to support an ergonomics investigation. These often include automatic (or some level of automation) support for calculation-based posture and manual handling assessment such as the NIOSH equation (Waters et al., 1993), RULA (McAtamney and Corlett, 1993), OWAS (Karhu et al., 1977) and REBA (Hignett and McAtamney, 2000). However, there are also DHM models which include calculation of energy requirements for a job based on input values such as gender, body weight, posture, cycle time and work time. Others can calculate the strength requirements of a variety of tasks based on either research into the strength capabilities of human beings or biomechanical models as discussed previously.

Other evaluation tools include vision analysis, which can include the generation of vision cones (either peripheral or foveal, and the ability to intersect these with the geometry of the workspace under investigation to determine limits of vision) or the ability to see through the manikin's eyes. Advanced vision features in some DHM packages include assessment of glare and reflections. Reach analysis can include the generation of reach zones from either acromion, full functional reach zone or convenient and comfortable reach.

Some DHM systems offer comfort assessment tools. These are most often based on joint angles in relation to the range of motion available for a joint (i.e. the notion that there is a more comfortable zone within the maximum range), forces exerted by muscles, pressure distribution on a surface area or repetition and duration of a task. Despite the developments in this area, there is a lack of empirical validation work comparing the predicted comfort ratings to those reported by real humans.

## ADVANTAGES AND DISADVANTAGES OF SIMULATION AND DHM

The applications of simulation and DHM are intrinsically linked to the advantages they offer. Thus, building on the previous sections, there are several arguments that can be made for simulation and DHM, which can be loosely grouped according to whether they relate to resources,



**TABLE 8.1**  
**Advantages of Simulation as a Human Factors Method**

| Resource Arguments                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Utility Arguments                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Safety Arguments                                                                                                                                                                                                                                                                                                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Reduced need for real objects/devices (e.g. physical prototype, vehicle). This is a critical issue for some high-cost machinery – for instance, the cost of a flight simulation is a fraction of the cost of using a real aircraft.</p> <p>Time/cost to build/develop – essentially, it is quicker to run an evaluation study using simulation compared to field settings.</p> <p>Cost of use (e.g. wear and tear, fuel for vehicles).</p> <p>May not need participants if using DHM approaches – potentially, a significant saving if needing to recruit (and pay) many participants.</p> <p>May not need facilities for testing if using DHM, thus saving in the cost of buildings, floor space, security, etc.</p> | <p><i>For designers/evaluators...</i></p> <p>Ask successive ‘what if’ questions – for example what if we moved a display from here to here – what would people be able to see?</p> <p>Adopt a more user-centred approach in which we start earlier in the design process and make more iterations of the design/evaluation cycle.</p> <p>Maintain confidentiality in design (with DHM), thus reducing the exposure of products to ‘real’ people.</p> <p>Target specific user groups, particularly limiting users who are either large or small on relevant body dimensions.</p> <p><i>For trainers....</i></p> <p>Possible to easily repeat exercises.</p> <p>Can easily assess an operator’s performance.</p> <p>Can give rapid and detailed feedback on relative/absolute performance, perhaps from different perspectives.</p> <p><i>For researchers...</i></p> <p>High control over the simulated environment.</p> <p>Performance measures can be easier to capture than in real world.</p> | <p>Reduce potential harm to people (users, investigators, patients, etc.).</p> <p>Consider rare, but critical events – such as a complication in a surgery, a pedestrian walking out in front of a driver or an engine failure on a plane.</p> <p>Consider issues for novice users (e.g. trainees) where knowledge/skills may be low, thus significantly increasing the probability of error.</p> |

utility or safety. Resources consider the costs in monetary value or time (or both) of prototypes, participants, exercises, etc. Utility arguments are related to the use or usefulness of simulation and DHM. Within this, a distinction can be made for different end users of the simulation, including designers and trainers. Safety arguments consider the well-being of users of a design, the investigator or the public (as patients, building occupants, etc.). Table 8.1 summarises these advantages, thus highlighting the breadth of positive reasons for using simulation in E/HF activities.

Considering *disadvantages*, perhaps the primary concern labelled at simulation relates to *validity*, that is the potential failure of the simulation to reflect real-world performance/behaviour as discussed. For ‘real’ humans using a simulation, their perception of risk, expectancies and/or motivations may differ from reality. For virtual humans in DHM, there may be issues with simplifying assumptions (accuracy) or a lack of completeness in the functionality of the software. With any simulation or DHM exercise, consideration must be given to the extent to which faith is placed in the results. For example, with high-fidelity simulators and photorealistic DHM models, it is easy to convince people that this is reality. From the authors’ experiences, colleagues have been quick to accept the outcomes of a DHM assessment, sometimes more so than results of a user trial with real humans! The manikins look convincing, and they have a 3D CAD form which can be manipulated

just like any other engineering component. Moreover, they simplify the variability of human behaviour and individual differences which can be difficult to conceptualise. Chaffin (2001) also raised concerns about designers' understanding of the limitations of DHM tools. Our responsibility as ergonomists is to convey these limitations to stakeholders and decision makers.

Other potential disadvantages of simulation are referred to in the literature (Duffy, 2009; Vincenzi et al., 2009) and will be specific to the context in which simulation is being used (equipment, tasks or environments). They include

- Potential for sickness, especially with immersive virtual-reality environments (see previous section)
- Lack of acceptance (by operators, trainees, designers or managers)
- Costs – both in start-up and maintenance
- Difficulties in using/learning to use software, which are often bespoke, low-volume products with bugs and a range of interface problems

## THE FUTURE

For the future of simulation, a range of developments can be envisaged, both in the short term and looking more medium or long term. The short-term outcomes that are relatively easy to predict are those based on linear extrapolations of existing trends, for example raising display resolution and increasing computer power and networking speed. As a result, it is likely that the visual fidelity of simulators will continue to rise, such that the visual experience of operators is indistinguishable from equivalent real-world scenery. Similarly, the appearance of DHM manikins will continue to improve, and processing time will continue to reduce. Current work on body motion tracking, partly a consequence of reductions in the cost of body motion tracking technologies, to support manikin positioning is likely to continue. This also provides opportunities to support dynamic assessments of work, which have traditionally required tedious joint-by-joint placements or selecting from libraries of pre-defined movement types. However, there remain challenges in matching the tracked human to the DHM manikin as well as issues of coverage of the entire human form (Colombo et al., 2012; Gragg et al., 2013).

The medium to long-term outcomes that are more difficult to predict will be based on recent developments and/or emerging areas. As an example, technological advances are providing greater realism in the simulation of non-visual senses, for example haptic and olfactory stimuli which may be included in simulators.

Future research opportunities also exist in modelling flesh and clothing deformation in DHM. This will increase the accuracy of clearance/fit investigations. Furthermore, integration of flesh deformation with the internal skeletal structure and organs may offer great predictions of human comfort or assessment of human injury (e.g. in handtool use). There are a range of other ergonomics issues which are currently inadequately addressed in the various DHM tools. These include environmental issues, the potential for slipping, emotions and psychosocial variables. The hand models are not always sufficiently detailed for the analysis of compact user interfaces, which are now ubiquitous in the consumer market. However, arguably, the next important step on the horizon will be the integration of cognitive models in the already well-established physical models.

## GUIDELINES FOR USING SIMULATION AND DHM

From the previously reviewed materials, we are able to generate guidelines for using simulation and DHM within E/HF activities. This is not a step-by-step guide for a simulation study, as this would be specific to the purpose of the investigation and context of use. Rather, it lists items which



should be reflected upon when using these tools to understand, evaluate or train human interaction with a product, workplace or system.

- Consider if simulation is the best approach. Before starting, evaluate the advantages and disadvantages of simulation/DHM and other approaches.
- Consider carefully which performance/behaviour measure you plan to take as well as the time required for analysis. Ensure that your investigation will provide the data you need.
- Understand what aspects of fidelity are critical to task behaviour/performance and develop or choose your simulator accordingly.
- Run a pilot. This is part of any good research investigation, but is particularly important when faced with issues of equipment/software reliability, as is often the case with advanced technologies.
- Provide a sufficient training/familiarisation period, so that you are certain that you are studying performance, rather than a participant's learning on how to use the simulator.
- Design the task, equipment, procedure and screening process to minimise simulator sickness. Have procedures in place should symptoms arise.
- Define a procedure for implementing DHM to reduce the impact of the assessor's subjectivity on the outcome.
- Consider how DHM is likely to be used within your design process – it cannot capture completely the rich and variable nature of human behaviour and performance, and therefore thought should always be given to testing with real users. This is particularly important when subjective data are required.
- Be cautious about adopting the default tall/short/average (sometimes 5th, 50th, 95th) percentile manikins. In many design cases, there is no value to designing for the 50th percentile as this provides no understanding on the range of people accommodated.
- Related to these, always do a sanity check when conducting a DHM activity (e.g. a simple measure of headroom (through the CAD system rather than the DHM software) and relate this back to anthropometric data) – it is all too easy to create an inaccurate manikin with a few erroneous mouse clicks, which may not be realised until later in the design process.
- Question the anthropometric data of any DHM used – consider nationality, age, clothing and secular growth.
- For both simulation and DHM, explain the limits of your analysis when presenting the results. Ensure the stakeholders and decision makers understand the validity of your work.

## CONCLUSIONS

It is clear that simulation, including DHM, has become an essential tool in the ergonomist's repertoire. A wide range of benefits can be described, and these, together with reducing costs and greater availability of the appropriate software/hardware, are leading to wider usage in E/HF research and practice.

Nevertheless, there are several prevailing issues for simulators, such as the requirements for fidelity, whether simulators are valid and how to minimise the sickness symptoms that can arise. For DHM, issues include how to incorporate dynamic movement and cognition in packages which have traditionally been used for static assessments of fit, reach and posture.

Research is being conducted to address these concerns, providing a better understanding of the experience of operators within varying simulation environments. However, there remains a fundamental recommendation to verify findings through real-world research and development, for instance, observations of drivers/pilots/medical staff, in the field. Moreover, further research work is needed to address the issues in defining the manikin's movement in DHM, and to develop and integrate models of behaviour and cognition to allow for a more holistic assessment of a user's interaction with a product and environment.

It is clear that simulation and DHM will continue to develop and be implemented in research and industry, helping to understand human performance and identify issues while avoiding risk at an early stage in the design process.

## REFERENCES

- Albers, K. and Abshire, K. (1998). Virtual human modeling for manufacturing and maintenance, SAE Technical Paper 981311, 1998, doi:10.4271/981311.
- Alexander, T. and Gunther, P. (2014). Ergonomic DHM systems – limitations and trends – a systematic literature review focused on the ‘future of ergonomics’. In *Third International Digital Human Modeling Symposium*, AIST, Tokyo, Japan, 20–22 May 2014.
- Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C. and Qin, Y. (2004). An integrated theory of the mind. *Psychological Review* 111(4): 1036–1060.
- Andrews, P.S., Polack, F.A.C., Sampson, A.T., Stepney, S. and Timmis, J. (2010). The CoSMoS process version 0.1: A process for the modelling and simulation of complex systems. Technical Report YCS-2010-453, Department of Computer Science, The University of York, York, U.K.
- Bach, K.M., Jæger, M.G., Skov, M.B. and Thomassen, N. G. (2008). Evaluating driver attention and driving behaviour: Comparing controlled driving and simulated driving. *Proceedings of 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction*, Liverpool, U.K. 1531514: British Computer Society, vol. 204.
- Baines, C.A. (2012). Virtual preparation to training for automotive assembly operations. MEng dissertation, The University of Nottingham, Nottingham, U.K.
- Banyard, P. and Grayson, A. (2000). *Introducing Psychological Research*, 2nd edn. Palgrave, Basingstoke, Hampshire, U.K.
- Beeney, B. and Charpland, J. (2009). Digital human modelling (DHM) in the automotive industry (Chapter 2). In N. Gkikas (Ed.), *Automotive Ergonomics: Driver-Vehicle Interaction*. CRC Press, Boca Raton, FL.
- Bowman, D. (2001). Using digital human modeling in a virtual heavy vehicle development environment. In D. Chaffin (Ed.), *Digital Human Modeling for Vehicle and Workplace Design*. Society of Automotive Engineers, Warrendale, PA, pp. 77–100.
- Bradwell, B. and Li, B. (2008). A tutorial on motion capture driven character animation. *Proceedings of the Eighth IASTED International Conference Visualization, Imaging, and Image Processing (VIIP 2008)*, Palma de Mallorca, Spain, 1–3 September 2008.
- Brolin, E., Högberg, D. and Hanson, L. (2012). Description of boundary case methodology for anthropometric diversity consideration. *International Journal of Human Factors Modelling and Simulation (IJHFMS)*, 3(2): 204–223.
- Bubb, H. and Fritzsche, F. (2009). A scientific perspective of digital human models: Past, present, future. In V.G. Duffy (Ed.), *Handbook of Digital Human Modeling*. CRC Press, Boca Raton, FL.
- Burnett, G.E., Irune, A. and Mowforth, A. (2007). Driving simulator and validity: How important is it to use real car cabins? In *Proceedings of Road Safety and Simulation Conference (RSS2007)*, Rome, Italy.
- Burnett, G.E., Large, D., Lawson, G., De-Kremer, S. and Skrypchuk, L. (2013). A comparison of resistive and capacitive touchscreens for use within vehicles, *Advances in Transportation Studies: An International Journal*, 31: 13–30.
- CAESAR (2006). Civilian American and European Surface Anthropometry Resource Project-CAESAR, available from [www.store.sae.org/caesar](http://www.store.sae.org/caesar) (Accessed 22 November 2014).
- Caon, M. (2004). Voxel-based computational models of real human anatomy: A review. *Radiation and Environmental Biophysics*, 42: 229–235.
- Chaffin, D.B. (2001). *Digital Human Modeling for Vehicle and Workplace Design*, SAE, Warrendale, PA.
- Chaffin, D.B. (2005). Improving digital human modelling for proactive ergonomics in design. *Ergonomics*, 48(5): 478–491.
- Chang, S.W. and Wang, M.J.J. (2007). Digital human modeling and workplace evaluation: Using an automotive assembly task as an example. *Human Factors and Ergonomics in Manufacturing & Service Industries. Special Issue: Advances in Agile Enterprises, Part I*, 17(5): 445–455.
- Columbo, G., REgazzoni, D., Rizzi, C. and De Vecchi, G. (2012). Industrial application of DHM for ergonomics driven by low cost mocap solutions. *International Conference on Innovative Design and Manufacturing*, Taipei, Taiwan, 12–14 December 2012.

- De Magestris, G., Micaelli, A., Evrard, P., Andriot, C., Savin, J., Gaudez, C. and Marsot, J. (2013). Dynamic control of DHM for ergonomic assessments. *International Journal of Industrial Ergonomics*, 43(2): 170–180.
- Demirel, H.O. and Duffy, V.G. (2007). Applications of human digital modeling in industry. In V.G. Duffy (Ed.), *Digital Human Modeling*. Springer-Verlag, Berlin, Germany.
- Du, J. and Duffy, V.G. (2007). A methodology for assessing industrial workstations using optical motion capture integrated with digital human models. *Occupational Ergonomics*, 7(1): 11–22.
- Duffy, V.G. (2009). *Handbook of Digital Human Modeling: Research for Applied Ergonomics and Human Factors Engineering*. CRC Press, Boca Raton, FL.
- Godil, A. and Ressler, S. (2009). Shape and size analysis and standards. In V.G. Duffy (Ed.), *Handbook of Digital Human Modeling*. CRC Press, Boca Raton, FL.
- Godley, S.T., Triggs, T.J. and Fildes, B.N. (2002). Driving simulator validation for speed research. *Accident Analysis and Prevention*, 34(5): 589–600.
- Gordon, C., Cg Churchill, T., Clauser, C.E., Bradtmiller, B., Mc Conville, J.T., Tebbetts, I. and Walker, R.A. (1989). Anthropometric Survey of U.S. Army Personnel Methods and Summary Statistics, Final Report (NATICK/TR-89/027) US Army Natick Research Development and Engineering Center, Natick, Massachusetts.
- Gragg, J., Cloutier, A. and Yang, J. (2013). Optimization-based posture reconstruction for digital human models. *Computers & Industrial Engineering*, 66(1): 125–132.
- Gwynne, S., Galea, E.R., Owen, M., Lawrence, P.J. and Filippidis, L. (1999). A review of methodologies used in the computer simulation of evacuation from the built environment. *Building and Environment*, 34(6): 741–749.
- Hanson, L., Blome, M., Dukic, T. and Hogberg, D. (2006). Guide and documentation system to support digital human modeling applications. *International Journal of Industrial Ergonomics*, 36(1): 17–24.
- Hanson, L., Hogberg, D. and Soderholm, M. (2012). Digital test assembly of truck parts with the IMMA-tool – an illustrative case. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41(1/2012): 2248–2252.
- Hignett, S. and McAtamney, L. (2000). Rapid entire body assessment (REBA). *Applied Ergonomics*, 31(2000): 201–205.
- Horgan, T.J. and Gilchrist, M.D. (2004). Influence of FE model variability in predicting brain motion and intracranial pressure changes in head impact simulations. *International Journal of Crashworthiness*, 9(4): 401–418.
- Howitt, D. and Cramer, D. (2011). *Introduction to Research Methods in Psychology*, 3rd edn. Pearson Education Ltd., Harlow, U.K.
- Karhu, O., Kansi, P. and Kuorinka, I. (1977). Correcting working postures in industry: A practical method for analysis. *Applied Ergonomics*, 8(4): 199–201.
- Karmakar, S., Pal, M.S., Majumdar, D. and Majumdar, D. (2012). Application of digital human modeling and simulation for vision analysis of pilots in a jet aircraft: A case study. *Work*, 41(2012): 3412–3418.
- Lamkull, D., Hanson, L. and Ortengren, R. (2009). A comparative study of digital human modelling simulation results and their outcomes in reality: A case study within manual assembly of automobiles. *International Journal of Industrial Ergonomics*, 39(2): 428–441.
- Landau, K. (2006). International Encyclopedia of Ergonomics and Human Factors. In W. Karwowski (Ed.), *Knowledge-Based Man-Modeling*, 2nd edn. CRC Press, Boca Raton.
- Laughery, R. (2005). Simulation and modelling for analysing human performance. In J.R. Wilson and N. Corlett (Eds.), *Evaluation of Human Work*, 3rd edn. Taylor & Francis Group, London, U.K., pp. 219–238.
- Ma, L., Zhang, W., Chablat, D., Bennis, F. and Guillaume, F. (2009). Multi-objective optimisation method for posture prediction and analysis with consideration of fatigue effect and its application case. *Computers & Industrial Engineering*, 57(4): 1235–1246.
- McAtamney, L. and Corlett, E.N. (1993). RULA: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24(2): 91–99.
- McDowell, M., Fryar, C.D., Ogden, C. and Flegal, K.M. (2009). Anthropometric reference data for children and adults: United States, 2003–2006. *Vital and Health Statistics*, 11(249): 1–68.
- Naghiyev, A., Sharples, S. and Yates, T. (2014). Train simulators for research. In M.S. Young and M.G. Lenné (Eds.), *Simulators for Transportation Human Factors: Research and Practice*. Ashgate Publishing Ltd. Farnham, Surrey, Canada.
- Nichols, S., Haldane, C. and Wilson, J.R. (2000). Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies*, 52(3): 471–492.
- Nichols, S. and Patel, H. (2002). Health and safety implications of virtual reality: A review of empirical evidence. *Applied Ergonomics*, 33(2002): 251–271.

- Nielsen, J. (1993). *Usability Engineering*. Academic Press, Cambridge, MA.
- Park, G.D., Allen, R.W., Fiorentino, D., Rosenthal, T.J. and Cook, M.L. (2006). Simulator sickness scores according to symptom susceptibility, age, and gender for an older driver assessment study. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, CA.
- Punte, P.A.J., Oudenhuijzen, A.J.K. and Hin, A.J.S. (2001). Ship bridge design and evaluation using human modeling systems and virtual environments. In D.B. Chaffin (Ed.), *Digital Human Modeling for Vehicle and Workplace Design*. SAE, Warrendale, PA, pp. 57–75.
- Rasmussen, J. (1987). The definition of human error taxonomy for technical system design. In J. Rasmussen, K. Duncan and J. Leplat (Eds.), *New Technology and Human Error*. Wiley, Chichester, U.K., pp. 23–30.
- Reason, J.T. (1990). *Human Error*. Cambridge University Press, Cambridge, U.K.
- Reason, J.T. and Brand, J.J. (1975). *Motion Sickness*. Academic Press, San Diego, CA.
- Reed, M.P. and Green, P.A. (1999). Comparison of driving performance on-road and in a low-cost simulator using a concurrent telephone dialing task. *Ergonomics*, 42(8): 1015–1037.
- Ren, A., Chen, C. and Luo, Y. (2008). Simulation of emergency evacuation in virtual reality. *Tsingua Science and Technology*, 13(5): 674–680.
- Reuding, T. and Meil, P. (2004). Predictive value of assessing vehicle interior design ergonomics in a virtual environment. *Journal of Computing and Information Science in Engineering*, 4(2): 109–113.
- Sabbah, A., Zaindl, A. and Bubbl, H. (2009). Design of a mock-up for supported ingress/egress using a DHM, SAE Technical Paper 2009-01-2268, 2009, doi:10.4271/2009-01-2268.
- Sharples, S., Burnett, G. and Cobb, S. (2014). Sickness in Virtual Reality. In B.K. Wiederhold and S. Banchard, *Advances in Virtual Reality and Anxiety Disorders*. Springer: New York, pp. 35–62.
- Stone, R.J. (2011). The (human) science of medical virtual learning environments. Invited paper. *Philosophical Transactions of the Royal Society B*, 366(1562): 276–285.
- Summerskill, S.J. and Marshall, R. (2011). Digital human modeling in the user-centered design process. In W. Karwowski, M.M. Soares and N.A. Stanton (Eds.), *Human Factors and Ergonomics in Consumer Product Design*. Taylor & Francis Group, New York, pp. 293–324.
- Vincenzi, D.A., Wise, J.A., Mouloua, M. and Hancock, P.A. (2009). *Human Factors in Simulation and Training*. CRC Press, Boca Raton, FL.
- Wang, Y., Zhang, W. and Salvendy, G. (2010). Effects of a simulation-based training intervention on novice drivers' hazard handling performance. *Traffic and Injury Prevention*, 11(1): 16–24.
- Waters, T.R., Putz-Anderson, V., Garg, A. and Fine, L.J. (1993). Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36(7): 749–776.
- Weiss, P. and Jessel, A.S. (1998). Virtual reality applications to work. *Work*, 11(3): 277–293.
- Wickens, C., Lee, J., Liu, Y. and Becker, S.E.G. (2004). *An Introduction to Human Factors Engineering*. Pearson/Prentice Hall, Upper Saddle River, NJ.
- Witmer, B. and Singer, M. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7: 225–240.
- Woldstad, J.C. (2006). Digital human models for ergonomics. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors*. Taylor & Francis Group, London, U.K., 2001; 2nd edn., CRC Press, Boca Raton, FL, 2006, pp. 3093–3096.
- Yan, X., Abdel-Aty, M., Radwan, E., Wang, X. and Cilakapati, P. (2008). Validating a driving simulator using surrogate safety measures. *Accident Analysis & Prevention*, 40(1): 274–288.
- Zehner, G. and Robinette, K. (2007). Use of CAESER data to develop multivariate cases and digital human models for the automotive industry. *Proceedings of the Second International WEAR Conference*, Banff, Alberta, Canada, 31 July–1 August 2007.

# *Section II*

---

## *Methods for Design of Work*

The second section of this book outlines a series of methods and approaches that are used to support design of work, systems, products and interfaces. The methods include those that have particularly evolved within ergonomics/human factors (E/HF) practice and those that have been adopted by E/HF from their use by those from other related disciplines. The methods within this chapter predominately adopt a systems approach and include tools that can be used through all stages of the systems life cycle.

This section builds on the overall methodological approaches presented in Section I and in the first chapter of this book, and considers the specific factors that need to be considered when applying them in particular contexts or with particular goals. The needs of a method and its output will inevitably be different if we are influencing the design of a complete system, compared with designing an individual system element.

The section begins with Chapter 9, by Houghton, Balfe and Wilson, which describes a range of approaches that can be taken to analysis and design of systems as a whole. They focus in particular on the now established concepts of human factors integration (HFI) and human systems integration (HSI) and emphasise the importance of a clear requirements specification in all elements of systems development. They also reflect upon how methods such as task analysis (previously discussed in Chapter 6) and allocation of function can be applied to systems design; there is a need here to retain well-established knowledge and approaches, but ensure that they are adopted in a flexible enough form to accommodate the current, complex nature of integrated, collaborative and dynamic work systems.

Chapter 10 provides a comprehensive overview of a set of methods that are used within user-centred design. They focus on describing not only different approaches to the design, but also the role of individuals within that design process, considering the skills required by different members of design teams. They note the distinct roles of the designer, the design researcher and the participant, each having valuable contributions to make to the design process, but none being infallible! They present an extensive list of methods and consider their usefulness for each of four stages in the design process. This chapter illustrates how far the discipline has moved beyond the previously dominant approach of user trials; whilst the authors of Chapter 10 demonstrate the value of this approach in the three later stages of the design process, they also emphasise the need for formative and need-driven approaches to provide a solid basis for design.



Chapter 11 presents approaches used to ensure that as wide a population as possible is accommodated within systems and product design. It links to the concept of accessibility and emphasises phenomena such as the aging population which, along with the increasing ages for retirement, is increasing the diversity of capabilities of those in many different types of workplaces. In addition to the traditional focus of E/HF on work, inclusive design is particularly relevant to consumer products, and many of the examples drawn upon in this chapter refer to consumer as well as work-related contexts.

Chapter 12 is the first of three chapters that consider a particular context in which we need to apply E/HF to design – warnings. Many of the general themes and tools described in Sections I and II are referenced, but the considerations associated with user behaviour in response to warnings, and how this can be taken into account in design, using a human information-processing approach, are presented. They note that current challenges for design of warnings in the twenty-first century include the increasingly diverse geographical and international nature of work (affecting varied work cultures, social cultures, cultural knowledge and beliefs and, of course, variation in languages spoken amongst a workforce). The authors note that the information-processing approach to design and analysis should not only lead to improved design but also support analysis of *why* a particular design is more or less effective.

In Chapter 13, Chris Baber notes that whilst E/HF and human–computer interaction (HCI) are distinct disciplines,\* there is much complementarity between their approaches. In the introduction to this chapter in the third edition of this volume, John Wilson notes ‘What is interesting is that the enormous concentration of human factors resources in HCI has allowed exploration of the use of virtually any method or technique – from psychological through to physiological measurement – and the combination of some of these into broad-based approaches to HCI evaluation’. What is important within the HCI community is that these methods are used with an appreciation of the underlying rationale and theory. Interaction with technology can of course range from virtual and augmented reality to smartphones to PC and tablet devices, and is becoming increasingly ubiquitous, with a ‘many-to-one’ (many computers to one person) picture emerging. In many cases, however, interaction with technology is not very ‘observable’ or overt, meaning that some traditional E/HF methods, such as structured observation, are very difficult if not impossible to apply. This is increasingly going to be a methodological challenge, as brain–computer interfaces come to fruition, and the methods we are currently using to evaluate work (such as the neuroergonomics methods described in Chapter 22) begin to merge with those which we may use to control our interaction.

The final chapter within Section II describes the way in which we implement methods in the context of control facilities design. Much ‘traditional’ ergonomics work was built on considerations of ‘controls and displays’; Chapter 14, by Charles, Sharples and Wilson, looks at the current trends in display and control technologies and presents the different stages of control facilities design, from initial analysis to integration and evaluation.

In all of these chapters, the responsibility of the E/HF practitioner remains to understand the strengths and weaknesses of the methods and tools in the varied contexts in which we might work and carefully select the approach which is most appropriate to the question at hand. An element that is particularly clear within this section is that not only will work contexts, tasks and goals differ, but so will users, and any methodological approach selected needs to take account of both.

**Sarah Sharples**

---

\* I would suggest perhaps more distinct than they should be.

---

# 9 Systems Analysis and Design

*Robert Houghton, Nora Balfe and John R. Wilson*

## CONTENTS

|                                                           |     |
|-----------------------------------------------------------|-----|
| Introduction.....                                         | 221 |
| Systems, Systems Engineering and Systems Ergonomics ..... | 222 |
| Systems Development Processes .....                       | 224 |
| Systems Development in the Real World .....               | 227 |
| Managing Human Factors in a Project.....                  | 230 |
| Human Factors Integration.....                            | 230 |
| Human Factors Scope .....                                 | 232 |
| Human Factors Integration Plan.....                       | 232 |
| Requirements .....                                        | 233 |
| Requirements Process .....                                | 235 |
| Representing Requirements .....                           | 235 |
| Writing an Ergonomics Requirement Document .....          | 237 |
| Evaluation of Requirements .....                          | 238 |
| Systems Analysis and Representation.....                  | 238 |
| Stakeholder Analysis .....                                | 238 |
| System Concepts and Views.....                            | 240 |
| Stories, Scenarios and Use Cases .....                    | 240 |
| Formal System Representations .....                       | 241 |
| Systems Analysis.....                                     | 241 |
| Function Analysis .....                                   | 241 |
| Task Analysis.....                                        | 243 |
| Allocation of Function .....                              | 244 |
| Final Thoughts .....                                      | 245 |
| References.....                                           | 246 |

## INTRODUCTION

Imagine that you are a software engineer, design engineer or architect. You have been given a leading role in a major systems design project – say a new airport cargo handling terminal or a new railway signalling centre. You have enough problems already, of course, but also you are aware enough to know that getting the human factors right will be vital and that you need to incorporate Ergonomics/Human Factors (E/HF) thinking from the start. But how should you do this? How will you integrate concerns for physical, cognitive and organisational ergonomics, and for the wide variety of stakeholders, who will use the system? How will you bring development of the human aspects of work design into line with development of hardware, software and organisational structures, and in what sequence will you do this?

Now, closer to home, imagine you are an ergonomist who is contacted by a company and asked to lead the human factors effort in a complete redesign of a factory facility or by a city metro operator who wants you to do the same for a city-wide transport system. Ergonomists often complain that they have not been brought in early enough in development, that people were considered too late in the design process and that as a result the E/HF interventions would be costly or the design



too ‘locked in’ to remove all its flaws. But coming in at the start with a blank piece of paper can be worrying, even frightening (e.g. Wilson and Whittington, 2001). Where do we start – with which of physical layout, task design, team selection and role design, training and awareness programmes, motivation and reward schemes, equipment and interface design? Each of these, and the design decisions made about it, will have implications for the others. Also, how do we know where we are heading – do we have a framework for the process or a roadmap?

The authors of this chapter have been in situations and development projects similar to those described earlier. These questions for E/HF – where do we start? What are the directions or options? How do we set priorities? How do we integrate human with technical, organisational and economic factors? When do we know if a design element is adequate? – are not easy to answer. They are important questions however, both in any particular project and also in general. To begin to address such questions, we need to understand the relevance of E/HF to the whole system life cycle, and especially to the systems development stages of specification, design and testing. In parallel, we need to understand the value and role of techniques of systems analysis and design in the application of E/HF. We also need to agree what we mean by a system and a systems approach.

This chapter opens by discussing the notion of systems in E/HF followed by a number of the better known specifications of the systems development process. We then describe in more detail the early, more analytical, stages in product and systems design, especially stages such as information gathering, requirements specification, function analysis and task analysis.

For much more information about systems analysis and design, and especially the role and contribution of E/HF, the reader should consult Chapanis (1996, p. 22), Hollnagel (2003), Meister (1985), Meister and Enderwick (2002), O’Brien and Charlton (1996), Vicente (1998) and Wickens et al. (1998).

## **SYSTEMS, SYSTEMS ENGINEERING AND SYSTEMS ERGONOMICS**

While there is an increasing consensus that, to paraphrase Hal Hendrick, ‘good ergonomics is systems ergonomics’, there is rather less certainty about what may actually be meant by systems ergonomics or, indeed, what constitutes a system at all. This is not merely a philosophical matter; how we think of systems will shape much of our involvement with their design and analysis. Within ergonomics, various definitions have been put forth that are worth briefly reviewing. An early view proposed by Singleton (1974) was that a system has related objects, undergoes changes over time and, in the case of a human-made system, has a purpose. Chapanis (1996) noted that the term system can be used in many different ways but concentrated primarily on ‘equipment systems’ defined as ‘an interacting combination, at any level of complexity, of people, materials, tools, machines, software, facilities and procedures designed to work together for some common purpose’. This view can be extended to consider that these interdependent components can be defined and properly understood only in relation to one another and their location within the whole (Luzeaux and Rualt, 2008, p. 12, quoting Sassure) and with reference to the goals of the system (Stasinopoulos et al., 2009). Systems may differ in terms of their complexity; simple systems can be easily described as organised set of sub-systems. More complex systems can be unorganised, making them difficult to describe in a structured manner, but their behaviour is essentially random and can be analysed statistically. Organised complex systems are the most difficult to analyse since they cannot be easily described and do not behave randomly; hence they are the focus of much of the systems engineering approaches (see Chapter 33).

Systems engineering itself also has difficulties in exact definition: Blanchard and Fabrycky (2011) suggest that we can however note, on the basis of different definitions that are in widespread use (p. 31), some common features: a top-down approach, a life cycle orientation, early concentration on defining systems requirements and an interdisciplinary/team-based approach to working. We will see throughout this chapter that these same qualities also apply to systems ergonomics in the context of systems development.

Wilson (2014) suggests that six notions can be used to explain the scope and interests of systems ergonomics:

1. *Systems orientation* – at the risk of stating the obvious, in systems ergonomics, the system itself is the primary focus of concern and the primary unit of analysis. It is important to move beyond the user-technology dyad and consider the wider system (or indeed, system-of-systems) in all its complexity. The may extend as far as considering how the system of interest interacts with other systems, either man-made or naturally occurring (e.g. the local geographical environment).
2. *Context* – we recognise that all work activity takes place within a context (e.g. physical or sociotechnical) which must be accounted for (see Moray, 1994). This provokes problems for the systems ergonomist in working out where the boundaries of analysis that define a system should lie and where the system gives way to sibling or infant systems. While there are no hard rules for doing this, it marks an early defining decision in any systems ergonomics venture.
3. *Interactions* – the nature of a system is that it consists of interacting parts that are defined in relation to each other and their interactions. In terms of systems ergonomics, this implies that we do not treat an item of equipment or a worker as if they were an island away from other workers, stakeholders, processes, etc.
4. *Holism* – a systems orientation implies that we see systems, people and artefacts as wholes thus consider users in the round as simultaneously physical, cognitive, social and cultural beings and similarly consider artefacts and processes in terms of multiple dimensions of safety, efficiency, reliability, performance etc.
5. *Emergence* – interactions produce complexity which can lead to the system exhibiting emergent properties as a result. Various forms of emergence can be observed in systems; for example, users not behaving expected, either in a way that is viewed as problematic (e.g. discovering safety rules cannot easily be adhered to in practice in the context of production process and pressures) or beneficial (e.g. overcoming difficulties in prescribed practices through improvisation, appropriating tools or procedures in ways not foreseen by designers and developing more effective procedures leveraging unforeseen local opportunities).
6. *Embedding* – the sixth feature of systems ergonomics is that good ergonomics E/HF is embedded within the practice of organisations and pursued in the field (which is to say, particularly in a large organisation, the ergonomics E/HF function itself is part of the wider work system).

On the basis of these notions, it is therefore proposed that systems E/HF can be defined thus:

Understanding the interactions between people and all other elements within a system, and designing in light of this understanding, a system being a set of inter-related or coupled activities or entities (hardware, software, buildings, spaces, communities and people) with a joint purpose; E/HF seeks to understand the links between the entities whether of state, form, function and causation; E/HF conceptualises any system of interest as existing within a boundary and thus a defined context, having inputs and outputs which may connect in many to many mappings; E/HF treats the system as holistic with the whole usually greater (more useful, powerful, functional etc) than the sum of its parts; and E/HF explicitly recognizes that the system changes and modifies its state and the inter-actions within it in the light of circumstances and events, thus showing emergent properties.

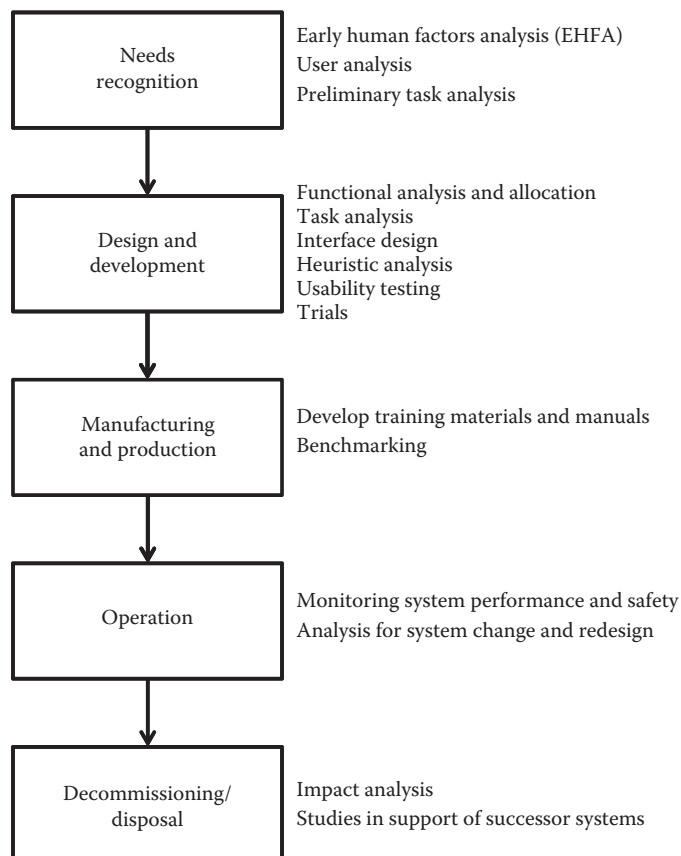
(Wilson, 2014, p. 13)

In terms of the example given of an ergonomist charged with undertaking the ergonomics on a complete redesign of a factory at the beginning of this chapter, this means that they would start by identifying the workers, supervisors and managers in the factory potentially affected by the changes, the equipment to be deployed and the location of the new factory. The processes and models presented in the remainder of this chapter would help the ergonomist define and analyse the factory as a system.

## SYSTEMS DEVELOPMENT PROCESSES

Although specific variants proliferate, it is possible to identify the simplified model in Figure 9.1 as a broad representative stereotype of a system development life cycle: an initial needs/objective phase is followed by the production of concepts/prototypes leading to a more formal design phase, manufacturing and production phases, an in-service phase and finally a disposal/recycling phase. Older models of systems development tended to concentrate only on activities leading up to product leaving the factory door, but in the modern era, it is more widely recognised that a product, system or service continues to be developed as it is used (e.g. maintenance activities, mid-life upgrades, adaptation and appropriation by users), and we also increasingly accept that designers have a responsibility to consider issues of disposal, decommissioning and recycling from the outset (e.g. see Ramani et al., 2010, for a review). As a result, it is more appropriate to think of a system in terms of having a development life cycle in which different stakeholders play different roles as it progresses.

This model is of course a very general stereotype, and domain- or industry-specific formulations may divide or package these stages differently and introduce evaluation gates or iterative cycles at different points, and so in practice, progress may not be so linear. Indeed, in design engineering, it is now widely recognised that strict adherence to a linear feed-forward interpretation of the systems development process is problematic in as far as it produces the ‘over-the-wall problem’ where seemingly plausible designs in terms of the requirements are batted back over the wall by manufacturing engineers as a result of parts that are physically impossible to produce or too costly to realistically produce, or production engineers ‘bat back’ manufacturing plans where the produced components



**FIGURE 9.1** Generic linear development process.

are difficult to assemble to cost and quality standards. The ‘Design for X’ movement (where X may be sustainability, manufacturing, assembly or maintenance, for example; see Huang, 1996) suggests that a more iterative, multidisciplinary approach to design ultimately pays off in terms of reduced delays, costs and quality issues. Ergonomists too should be mindful of the consequences of early design decisions in terms of their later implications in the life cycle of a product or system (e.g. training needs for system maintainers will be influenced by decisions made at the initial point of design) and that at each stage in that life cycle, different ergonomics activities may become required.

The earlier generic model can be found at different scales too. In smaller projects, perhaps carried out by an SME, the phases may undergo some compression, but it is also recognisably similar to the CADMID cycle (concept, assessment, demonstration, manufacture, in-service, disposal) used by the U.K. Ministry of Defence to manage the procurement of major platforms (Figure 9.2).

Well-known variants on the linear process come from software engineering, and this has become known as a waterfall model, in which each activity leads naturally into the next (Figure 9.3) albeit with iterative cycles made explicit.

In more recent contributions, specialists in software engineering, and particularly those involved in interface design, have expanded upon this waterfall model, to define a V model (in which the two stages of the cycle are structured into two processes, downward for specification design and upward for validation and testing); a spiral model (e.g. Figure 9.4), in which we start to see representation of the iterations that should take place in development, where needs are formulated progressively, risks analysed and resolved as and when they are encountered; and an incremental model, which also explicitly recognises the iterative development of a design (see Lepreux et al., 2003).

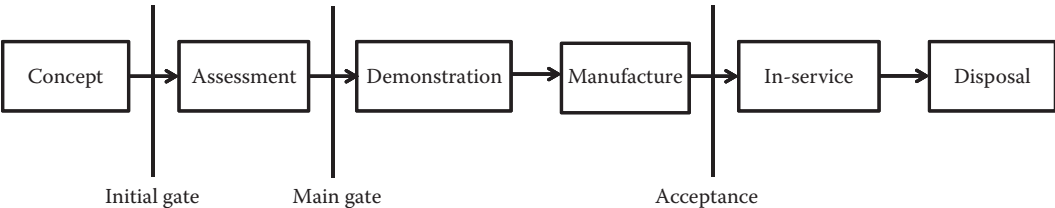


FIGURE 9.2 The CADMID procurement cycle.

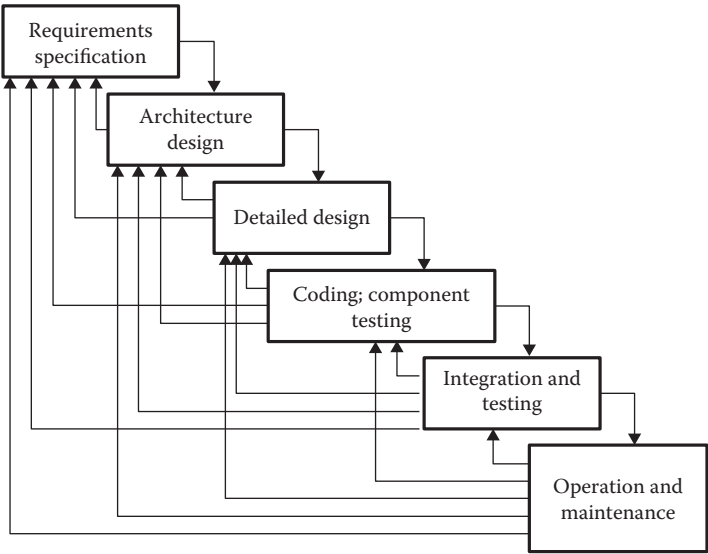
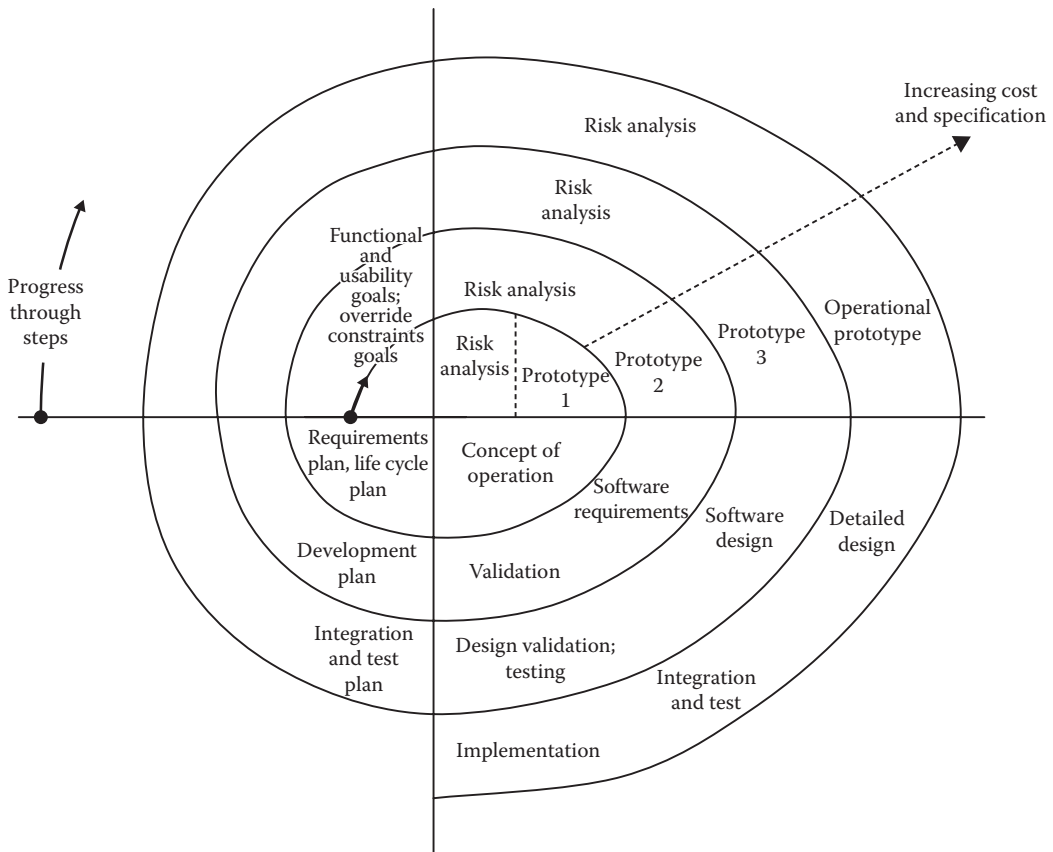


FIGURE 9.3 The waterfall model of the software life cycle.



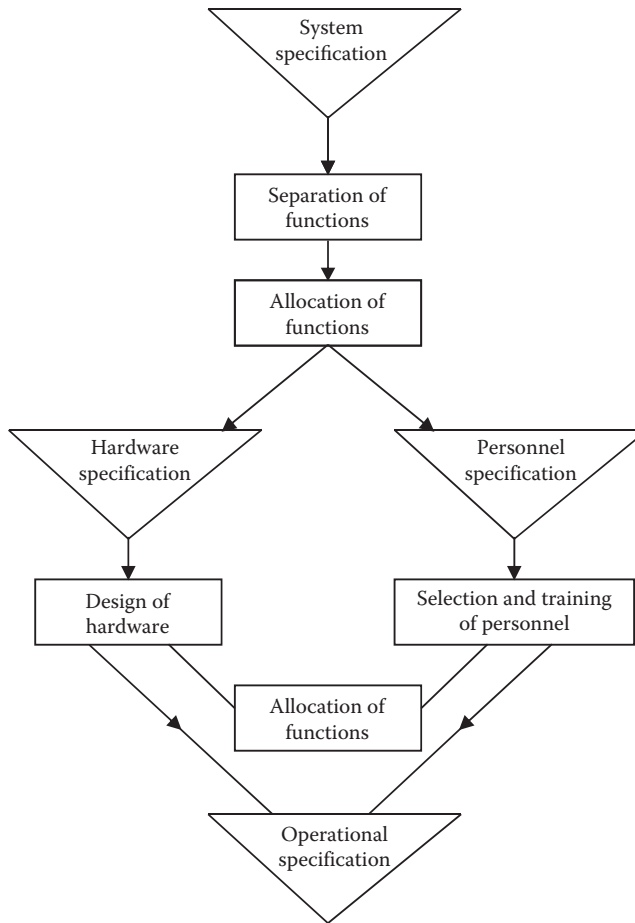
**FIGURE 9.4** Spiral model of design.

Early proposed processes for man-machine (nowadays human-machine) systems design were similar to that of Singleton (1974) whereby human and hardware subsystems are developed in parallel, followed by subsequent integration in terms of both the interface and then the operational system itself (Figure 9.5). It is probably no accident that models of this kind were proposed at a time of great interest in socio-technical systems design (e.g. see Cherns, 1987; Clegg, 2000), with its emphasis on joint development of social and technical systems, although it has to be said that practice did not always match theory.

Subsequently, representations of the systems design process have tended to handle the parallel hardware and human design needs by binding ergonomics consideration to stages in the system's overall design process (e.g. Bailey, 1989; Kirwan and Ainsworth, 1992, p. 28; Kragt, 1992, p. 15; Sanders and McCormick, 1987, p. 552; Salmon et al., 2010, section 36–31; Wickens et al., 1998). Representative types of activity are noted in Figure 9.1. Alternatives to this emphasise a more explicit parallel process of the development of jobs, machines and interfaces (Kragt, 1992, see Figure 9.6; Aikin et al., 1994; see Figure 9.7).

Whatever the development process used, it is possible to identify some common activities that ergonomists will need to carry out at all stages.

- Work from the 'what' to the 'how', from what the functions, roles and responsibilities should be and where and in what context these should take place, through to what controls, displays, equipment and workplaces needed to do this.
- Jointly consider physical, cognitive, psychological and social interactions.



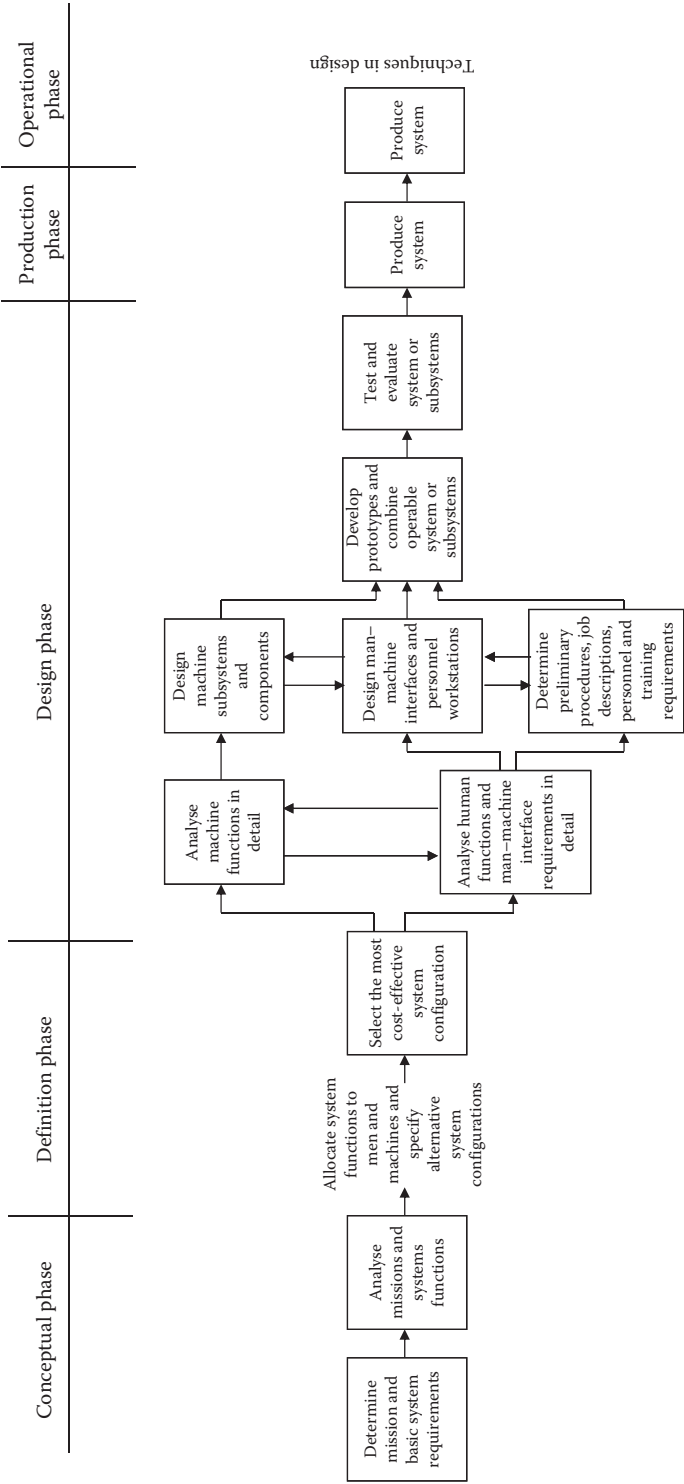
**FIGURE 9.5** The systems designs process. (Adapted from Singleton, W.T., *Man-Machine Systems*, Penguin, Harmondsworth, U.K., 1974.)

- Support design for the organisation of work, for jobs, roles and teams, as well as for hardware, software and workplaces.
- Provide evidence and source traceability, including rationale, for all requirements and analyses generated.
- Be prepared to negotiate and if necessary assert the importance of considering ergonomics E/HF within system design.

## SYSTEMS DEVELOPMENT IN THE REAL WORLD

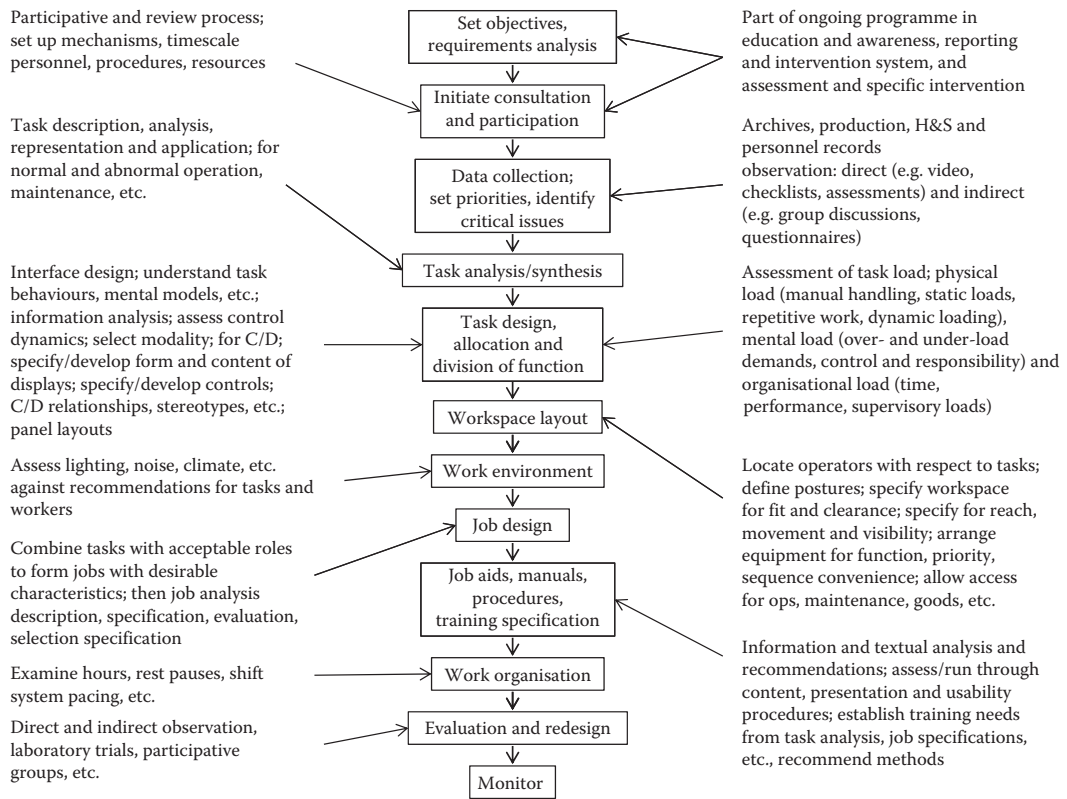
It is somewhat ironic that so much effort has gone into so many descriptions and specifications of the design process, since design in the real world so rarely goes neatly to plan. Anyone who has been involved in a design, whether large or small, will know that it rarely if ever follows the pre-defined path and that development will actually include the adaptation of planned stages, iterations around loops in the process, frequently redefined objectives or means, and ever-changing priorities.

In real developments, rather than a neat agreement on goals and then setting of requirements and constraints early on, we will usually find several competing requirements and goals. Classically, marketing may require one thing – something very eye catching for the customer with a large number of variants to allow targeting of many markets – whereas manufacturing would prefer something which requires minimal re-settings of machines, no difficult machining



**FIGURE 9.6** Design activities occurring during phases of the system development process. (Adapted from Kragt, H., *Enhancing Industrial Performance*, Taylor and Francis, London, U.K., 1992.)





**FIGURE 9.7** Ergonomic design process. Each of the stages of ergonomics in itself has particular input requirements, which imply use of a large variety of methods in analysis, design and evaluation.

processes and with as few variants as possible. In another example, some parts of a service company may think it best that their new computer system be merely an adaptation and improvement of the one they already have and which is on the market; others in the company, perhaps more adventurous, may be pushing for a radical rethink and an entirely new form of system. In complete contrast, in other development processes, the team may be bereft of early ideas. They may know that they need to have a new product for the market but may not know where to start or may lack inspiration.

Whether we have many competing requirements or a paucity of ideas, we are not helped when there are considerable gaps in the information and market intelligence available to the company, or when the information available is contradictory and confusing. Much design, in fact, does not start from a clean new set of goals and ideas for completely new design, but consists of development teams begging, borrowing or stealing ideas from others. One large U.K. domestic products manufacturer known to the author in fact admitted, in confidence, that the majority of their development started from visiting trade fairs overseas and seeing what other companies were putting on the market.

Once the design process actually starts, again the process does not necessarily match the classic or idealised system, of a neat move from conceptual design to various levels of detail design, albeit with iterative steps. In fact, there may be a great deal of developing and trying out various designs of different levels of sophistication, prioritising these in a crude fashion as the team goes along. When this is working well, as the models or prototypes themselves increase in sophistication so can the test and evaluation methods, including those for E/HF, become more detailed and extensive.

In practice, as the design process goes on, and for anything which has any substantial amount of resource invested, there will be pressure to short cut. This pressure is usually from the finance or

project management department, but may also be from marketing, who want to get the product to the customer as soon as possible, and perhaps also from production who want to see their facilities being fully employed. It is often a fact of life that prototypes are often retained as a result of time/financial pressures and 'made to work' rather than as, would be the platonic case, disposed of as inadequate and the lessons learned applied to another fresh prototype. Also of critical importance to ergonomics, there may often be a late realisation that the real user needs and requirements have been forgotten, leading eventually to system misuse or disuse, or at the least expensive reworking and modification at a late stage in design. A strong Human Factors Integration (HFI) programme of work may help address these concerns by placing an emphasis on understanding user needs in the system and guiding the design from the earliest possible stage.

## MANAGING HUMAN FACTORS IN A PROJECT

### HUMAN FACTORS INTEGRATION

There are many ways in which human factors issues in a system might be identified; this book provides many different approaches and systems of classification. In terms of how this is structured in large projects, particularly in the oil/gas and defence sectors, the notion of HFI (sometimes referred to as human systems integration, HSI) has become popular and, in some cases, mandatory. HFI/HSI is the current incarnation of a process to reconcile the top-down nature of systems development with the bottom-up nature of user-centred design approaches that began with the US Army's MANPRINT initiative (for a comprehensive review, see Booher, 2003). There are six core domains of HFI which represent different sets of human-centric concerns; manpower, personnel, training, human factors engineering, health hazard assessment and systems safety (Table 9.1).

In the original MANPRINT formulation, a seventh domain, that of solder survivability, was also present although this may be substituted with habitability or in the case of the U.K. MoD, a social and organisational domain (Forrest and Fletcher, 2009). Each domain raises a set of issues to be considered across the life cycle of a system and also forms a venue for convening relevant stakeholders. For example, in the case of manpower, we might ask how many people are required to operate a system and bring together operations leaders, recruiters, end users and planners to understand this. The answers may not be immediately intuitive either; for example while a skeleton crew might be able to operate the semi-automated systems of a modern warship, a somewhat larger crew would in fact probably be required to maintain a baseline level of hygiene and maintenance. In defence settings specifically, HFI will be the responsibility of a HFI lead working as part of an integrated project team alongside specialist engineers and other parties.

The venture of carrying out HFI can be difficult in a range of ways; in particular, the HFI lead may find themselves challenged in terms of exerting leverage over design decisions and influencing the direction of the overall project. The wider issues of asserting the value of HF that would be recognisable to past generations of ergonomists may also be present (e.g. 'it's all common sense' and the sense that HF is in some ways out of step with other engineering disciplines whose members have more of a shared educational background). While this remains an ongoing challenge for our discipline, there are some ways in which these gaps can be dealt with. First, one approach may be to attempt, where possible, to use forms of representation or convert HF representations into more widely recognised forms of representation to aid understanding within the team (e.g. Houghton, Baber and Chaudemanche, 2008; Houghton, Baber, Cowton, Walker and Stanton, 2008). If E/HF is part of the systems engineering function, then we need to be able to communicate successfully with the other members of the team. Ergonomists, in the main, use techniques and methods whose outputs are not mainstream and not always easily assimilated by non-ergonomists. That said, we should also be mindful of the risk that important information or even core values may get lost in the process. For example, the authors are aware of incidences where pressure has

**TABLE 9.1**  
**Human Factors Integration Domains**

| Domain                    | Issue                                                                                                             | Topics to Consider                                                                                                                                                                                                                                                                                     |
|---------------------------|-------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Staffing                  | How many people are required to operate and maintain the system?                                                  | Staffing levels<br>Workload<br>Team organisation<br>Job specifications                                                                                                                                                                                                                                 |
| Personnel                 | What are the aptitudes, experience and other human characteristics necessary to operate and maintain the system?  | Selection, recruitment and career development<br>Qualifications and experience required<br>General characteristics (e.g. body size, strength, eyesight)                                                                                                                                                |
| Training                  | How to develop and maintain the requisite knowledge, skills and abilities to operate and maintain the system?     | Identifying the requirements for new skills<br>Documentation<br>Training courses<br>Requirements for specialist training facilities<br>Individual and team training<br>Skill maintenance (e.g. refresher courses, drills)                                                                              |
| Human Factors Engineering | How to integrate human characteristics into system design to optimise performance within the human/machine system | Equipment design<br>Workstation/console design<br>Workplace layout<br>Maintenance access and ease of maintenance<br>User interface design (e.g. computing facilities and screen design)<br>Function allocation (between humans and automation)<br>Working environments (e.g. climate, lighting, noise) |
| Health hazards            | What are the short- or long-term health hazards resulting from normal operation of the system?                    | Exposure to toxic materials, electric shock, mechanical injury, musculoskeletal injury, extreme temperatures, optical hazards, electromagnetic radiation                                                                                                                                               |
| System safety             | How to avoid the safety risks that humans might cause by operating or maintaining the system abnormally?          | Sources of human error<br>Effects of misuse or abuse<br>External and environmental hazards                                                                                                                                                                                                             |

*Source:* Adapted from Widdowson and Carr (2002), p. 4.

been put on ergonomists to generate probabilities around overly complex human error issues for inclusion in common system risk assessments where a more responsible and useful product would be to emphasise the sociotechnical complexity associated with a specific human risk and, arguably, exactly why its nature makes it irreducible to a single numerical statement of probability and severity. Second, suggestions can also be made to overcome the ‘soft’ organisational and social issues through improving the profile of E/HF specialists, improving communications, demonstrating the benefits of E/HF to engineering design and even having ergonomists undertake further engineering training to help them understand the viewpoints of other groups (for a larger review and framework, see Waterson and Kolose, 2010). Finally, HFI should by nature integrate with the programmes of work planned by other disciplines on the project. This involves understanding the goals, approaches and timescales for each discipline and identifying where and how value can be added through ergonomics analysis. Aligning with project plans and goals can help achieve acceptance and understanding from engineers who have not previously interacted with E/HF specialists.

## HUMAN FACTORS SCOPE

In an ideal situation, ergonomists would have been integral to the initiation of a project and present throughout the definition of the concepts behind the system. However, in the real world, this is often not the case, and thus the first contribution, having understood the project or design brief, is the human factors scope. This is the identification of E/HF impacts and issues and provides a human factors roadmap for the wider project. It means identifying which aspects of the system and its operation have an effect on the well-being and performance of people and what we do (or should) know about people's characteristics and behaviour which might have an impact on successful system performance. In some settings, this may more commonly be referred to as carrying out an Early Human Factors Analysis (EHFA) with the identified issues added to a Human Factors Issues Register cross-referenced against design and, where appropriate, the project's overall Project and Safety Risk Registers.

## HUMAN FACTORS INTEGRATION PLAN

A first human factors integration plan (HFIP) is often developed for larger projects at this time, to an extent drawing upon or formalising the human factors process requirements in the light of the HF scope and initial analyses. The HFIP defines the integration of ergonomics into systems development and also defines relevant assurance procedures to monitor and sign off activities required to support such integration. It ensures that knowledge of people's capabilities and characteristics is fed into the system life cycle at the best time for a successful and efficient development process. HFI and HFIPs may be developed and applied to particular projects or systems designs but can also be a generic document that can be operationalised for any particular project or system. Some more complex projects may ideally require the HFIP to be updated to reflect changes in the design philosophy, project phase and/or the project structure. One of the tensions around the generation of HFIPs is that within the structure of some projects where it takes the form of an agreement between the ergonomists and the project management team, it provides the best leverage ergonomists have over influencing the system or product at hand over the course of its development and use. Consequently, it may have an unfortunate tendency to bloat into a weighty document consisting of many hundreds (if not thousands!) of pages covering every conceivable eventuality and 'pouring concrete into the veins of creative design'. Therefore, care should be taken to make the level of detail appropriate to the project size and its good regulation.

The content of an HFIP will usually include

1. Identification of those responsible for particular aspects of the human factors systems development and the arrangements made organisationally including lines of reporting.
2. Providing for end-user representation, how stakeholders will be consulted and preferably enabled to participate in the project or development.
3. Showing how there will be continuation of availability of expertise, how the target company and suppliers will have human factors expertise throughout the system development life cycle.
4. Coordination of human factors delivery across all the organisations involved in the project, including subcontractors, and how decisions will be made, priorities set and unresolved issues settled.
5. Understanding relevant regulations for the project and how human factors considerations will be incorporated into the consultation process with the regulators.
6. Identification of operational concepts, their content, and how and when these will be produced.
7. The analyses required of any existing or related system (technical or organisational) in order to identify and assess key aspects of people, processes, equipment, performance and tasks.
8. Definitions of measurement methods and criteria in order that outcomes of performance and well-being can be measured.

9. Listing of relevant human factors standards and guidelines, from the particular domain and from elsewhere.
10. The management of contractual arrangements so that these fit with the requirements of the overall project.
11. Definition of detailed requirements for human performance, including reliability, consequent loading on people (including workload and situation awareness) and risk of injury or ill health.
12. Detailed criteria and methods for the evaluation of the operability and performance of the total system and its design.
13. A timeline plan for formative and summative evaluation of the whole system and any coherent sub-systems.
14. Support requirements for documentation, procedures and training throughout the project.
15. Initial plans for human factors testing of the system prior to and during the commissioning process.
16. Provision for monitoring of the performance of the system once it has been commissioned and is in full operation.
17. Full description of terms and abbreviations used.
18. Reference to further sources useful to the different parties using and drawing from the HFIP.

The early human factors content specified in an HFIP will include: user tasks and responsibilities; user characteristics and competences needed; interactions within and between roles; information requirements and support; workstation assets; operability targets; workload assessment and staffing levels; training needs analysis; functional allocation (including automation); job designs; workstation options; site layout (building, open space, control room, etc.); user interface designs and user support needs.

The HFIP should also include some specification of the human factors content of the project. This would include topics related to the understanding of the users, their tasks and responsibilities, their characteristics and competences, job design, as well as the interactions between roles. Analysis of the technical system might include information requirements specification, workstation assessment, site layout options and assessment. The interaction between the staff and the technical components might drive workload assessment, training needs analysis, function allocation and user interface design and assessment.

## REQUIREMENTS

It is difficult to overstate the vital role that user needs and E/HF requirements play in systems development. To ignore them, misunderstand them or identify them carelessly invites potential for systems failure and wasted resources of time, people and costs. To quote from the National Audit Office report into the failure of the U.K. FIREControl IT project, 'FIREControl was flawed from the outset because it did not have the support of the majority of those essential to its success—its users [p. 5]... there was no single, authoritative owner of the user requirements...' [p. 6] (National Audit Office, 2011). Lindgaard et al. (2006) identify other 'spectacular' systems failures which they say highlight the vital importance of the requirements capture phase and that the leading reason for the failures was that 'project teams did not know how adequately to generate user information requirements' (p. 50). They go on to say that the motivation behind requirements engineering is to reduce the high cost of misunderstanding between users and designers so that the systems are built to do what users want, on time and at an affordable cost. Referring to a 2006 report on technology project failure, Pew and Mavor (2007) found that only 34% were successful, and five of the eight most frequently cited causes of failure were associated with poor user requirements (incomplete requirements, lack of user involvement, unrealistic user expectations, requirements continually changing system no longer needed).

Systems specifications in general, and ergonomics and user requirements in particular, walk a tightrope between generality and specificity. If they are too general, then they will seem applicable to all systems and circumstances but actually be of little practical assistance to any one project. If they are too specific to a context or type of system, then they will rarely be transferable to other contexts or systems, meaning at the least inefficiency in effort and at the worst apparent incompatibilities between ergonomics requirements for what appear to be similar projects at different times.

There are different classes and forms of ergonomics requirement. No one is necessarily better than another, and all can have their place in supporting systems development. However, ergonomists, the engineers, developers and business managers they work with, and the clients must all be quite clear about which sort of ergonomics requirement is being produced and discussed. This clarification is often aided by ensuring that different classes of requirements are produced and collected together, but not mixing up forms and classes (although classes may overlap). Drawing from the discussion earlier, classes and levels of ergonomics requirement include the following:

*Ergonomics requirements* – for all the relevant human factors of the development and implementation, including the human–machine interface, workplace and work environment, jobs and organisation of work, support through training and supervision, maintenance, etc. Broadened from the user requirements by inclusion of concerns from the wider socio-technical system, and modified through filtering, feasibility testing and assessment against good ergonomics knowledge and practice.

*Ergonomics process requirements* – produced early on for a project and related to the HFIP where one exists, defining how the ergonomics should be incorporated into the project.

*User requirements* – drawn directly from user needs and stakeholder analyses and scenarios, relevant to endusers, customers and all others potentially effected, and generally focused on the human–machine interaction and immediate workspace and job support.

*Usability requirements* – generally concerned with human–machine interaction and should be included within the ergonomics requirements, addressing, effectiveness, efficiency, utility, accessibility, satisfaction and so on.

*‘Look and feel’ requirements* (Robertson and Robertson, 2006, p. 176) – describe the intended spirit, the mood or the style of the system’s appearance, and are related to ideas of emotional design in human factors (these are sometimes seen as a part of usability definitions and requirements).

*Guidelines and standards requirements* – may be included wholesale, referred to within the ergonomics requirements, or tabled separately. Care should be taken over which standards are truly relevant so as not to require too much from the developers.

*Human impacts requirements* – how the system should support user health, satisfaction, comfort, etc.

*Human–machine system performance requirements* – defines what people must be capable (and supported) to do, and what the human–machine system as a whole must achieve, often in quantifiable terms if these are available, and may include target time and error rates, recovery times and rates, outcome quality, costs, etc.

*Ergonomics and environmental sustainability requirements* – these are relatively recent additions to the range and refer to ergonomics sustainability (participation, health and safety, adaptability and flexibility of the work systems, etc.) and the human factors–oriented issues of environmental sustainability (power efficient performance [routing trains], use of materials in engineering, minimising environmental impacts such as noise and lighting, etc.).

*Implementation requirements* – user support, learning and training needs, monitoring and continuous improvement; in an ideal process, these are integrated with usability requirements in the Ergonomics Requirement Document (ERD).

Requirements may be identified directly from user or function analyses, but may also be derived from the implications of other higher-level requirements.



## REQUIREMENTS PROCESS

It has already been made clear that the process of generating, and the actual content, of ergonomics requirements will vary according to the size and complexity of the system being designed and the resources available. The level of effort required for a major renewals project of several years' duration will be very different to that for a single small piece of technology for a specific task.

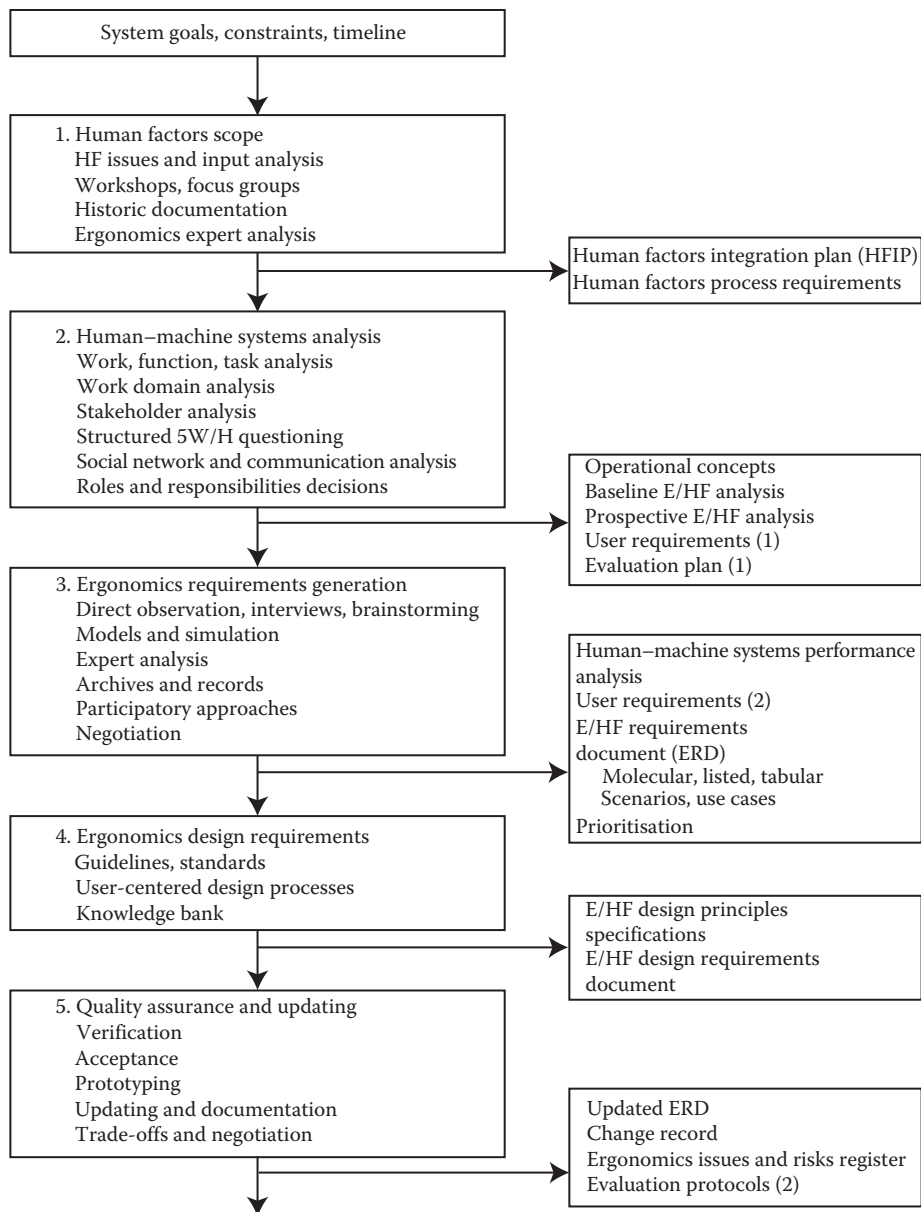
It can be tempting for ergonomists to try to include everything identified as relevant to people and human performance within the E/HF requirements. However, all agreed requirements will have to be accounted for by developers, suppliers and the procurement process, perhaps governed by contractual terms, meaning that any proposed system or system component will have to be assessed or tested against the requirements subsequently. It does not help efficient development or the credibility of ergonomics if this assessment and agreement process takes too long or is too difficult to carry out for requirements which could be reasonably considered as optional. Therefore, the process and stages of ergonomics requirements through the system life cycle must be set to be appropriate to the project importance and complexity, system size and the design focus – not everything in this chapter will be needed or even advisable for every project. There are a number of frameworks proposed for the requirements process, one of the best known being Volere (Robertson and Robertson, 2006). This is a quite detailed multistage process, and a more straightforward process of developing ergonomics requirements will follow the stages summarised in Figure 9.8 and described in the rest of this section and consists of five key stages.

After understanding the project or system goals and constraints, first, the E/HF scope must be set, highlighting what human factors are important for the particular target system. Second, a human-machine systems analysis is carried out – assessing and understanding more about the key human factors in the current system (baseline) or work domain and identifying the critical factors for the future system (prospective) and the impact of these. Stakeholder analyses are also carried out now. Third, the E/HF requirements are generated and delivered in an ERD, based on user needs, user requirements and human factors performance requirements (what must the human-machine system achieve and what are the acceptable and unacceptable impacts on the users and other stakeholders). Fourth, the E/HF design requirements for the system are defined – what are the people-centred design priorities, what form must the human-machine system design take and how will its ergonomics be evaluated? This stage will take place only for some systems and projects. Fifth, all requirements must be quality assured and validated, and for any large system, they will need to be assessed continually and updated where necessary. For smaller and less complex system designs, the process may comprise only certain stages and will certainly use only a subset of the analysis techniques and sources of information available at each stage. For larger systems, there are likely to be more loops of iteration, returning to the previous stage for more insight in the light of issues raised at the next stages. However, even for larger systems, not all the techniques and methods described in the following sections will or even can be used. They are summarised here to alert readers to alternative ways to get to key user needs and E/HF requirements, from which the ergonomist will select as appropriate. Therefore, all the inputs, activities and outputs outlined later in this chapter are indicative, not prescriptive, and give a general idea of the way in which ergonomists will approach developing ergonomics requirements in systems design so as to enhance the human factors of the operational system.

## REPRESENTING REQUIREMENTS

The most common form of representation is textual sets of requirements, presented as a simple structured list, in a tabular format or distributed on a series of 'shell' cards (postcards – Robertson and Robertson, 2006). There are a number of reasons: they are the easiest to generate and present; they are the only feasible way of writing very large numbers (hundreds and thousands in some cases) of requirements; hierarchies of requirements can be represented; they can often be linked





**FIGURE 9.8** A process for the development of human factors requirements.

directly from a hierarchical task analysis or function analysis (see Chapter 6); they are the easiest form to which to allocate priorities; and perhaps most importantly, they provide the simplest way of checking off supplier and design conformance as development progresses. They are sometimes called molecular or atomic requirements as they represent what is required in detail and many statements. However, there are other forms of representation, mostly useful either in the very early stages of large projects or in user-centred design for smaller projects and less complex products or systems. These other forms are basically scenarios, but may sometimes be labelled as stories and use cases as well as scenarios. Presentational forms of scenarios may be text (literally stories), graphical, pictorial, computer-based multimedia or virtual reality models, and hard models. Whatever form of representation is used, user requirements and ergonomics requirements must be

provided in a systematic, structured and usable fashion. Each requirement must have an identifiable source, and rationale, interactions and dependencies between requirements must be defined. Any E/HF Requirements Document, especially built around lists or tables (but to an extent relevant to other approaches), should start by identifying

- The system being developed and the design stage if this is relevant
- Project or design goals and any translation of these into ergonomics terms
- Critical constraints on design and ergonomics
- End users and other stakeholders
- Any assumptions made

Good sets of requirements should be

- Relevant to the focus system
- Generally independent of the existing system
- Accurate in how they draw upon ergonomics knowledge
- Complete but not so detailed that they become a block to development
- Consistent within themselves and with earlier requirements sets for similar systems
- As unambiguous as possible requiring minimal reinterpretation
- As complete as possible or at least with known gaps identified
- Verifiable with traceable sources
- Prioritised
- Integrated within the development process, design methods and procurement process

However, these guidelines are just that and may be not followed in certain circumstances depending on the project, system or process; for instance, there may be good reasons to build requirements around an existing system, or for them not to be prioritised at certain stages. Ergonomists need to help the project to understand early on that requirements, and therefore specifications to do with E/HF will sometimes of necessity be ambiguous; ergonomics terms like easy to use or user compatible are too vague to be of help without further specification. Other terms and notions such as mental model, mental workload and situation awareness are increasingly employed by a variety of designers and engineers as well as ergonomists, often without a clear idea of what is meant and, at best, may have multiple meanings for different people and at worst may be interpreted in a ‘common sense’ way that causes confusion. The requirements in their most detailed form, used for highly critical developments, would include a unique reference number, an identifier to link them to system facets, a description, rationale, source(s), the criterion met, likely customer or project value, dependencies, conflicts, supporting materials, issue classification (performance, systems safety, health and safety, compliance, assumptions, project risks), prioritisation and a history of changes made. Of these, even the lightest touch sets of requirements must include a unique reference number, a description, rationale, source(s), conflicts, supporting materials, prioritisation and a history of changes made.

## **WRITING AN ERGONOMICS REQUIREMENT DOCUMENT**

There are a number of guidelines which will help the writing of an (molecular, tabulated or listed) ERD. These are relevant particularly to list or tabular sets but also to tabular scenario-based requirements (see also FAA, 2003):

- a. Be as brief and succinct as possible in each requirement, referring the user/client to other information sources only where appropriate.
- b. However, do refer to previously agreed standards, guidelines and processes within the requirements where this will simplify and shorten the content.

- c. Where other documents are referred to, then it should be made clear which parts of these standards or guidance are NOT relevant to the system under consideration.
- d. Other than (b) and (c), limit the number of different reference documents used in the ERD to avoid adding complexity.
- e. Be as precise and specific as possible so that the requirement can be addressed during test and evaluation (especially for the detailed or final design ergonomics requirements).
- f. Where possible, define and operationalise a requirement to be relevant to the particular stakeholders, tasks and contexts of use rather than using general requirements.
- g. Specify or refer to expected human-system performance levels where possible and useful.
- h. Include a descriptive non-binding rationale or other background for every requirement, especially where the requirement is non-specific or requires explanation.
- i. If the system interfaces with other (new or existing) systems, ensure that the requirements are compliant with standards or guidelines for those systems and identify the interfaces with the relevant requirements.
- j. For listed or tabulated sets, use one requirement statement or scenario per paragraph number.
- k. In writing a requirement statement, employ a subject–predicate format with simple sentences and with no compound predicates – for example the xxx shall be yyy.
- l. The list or table of requirements, or set of scenarios, should include a clear set of priority ratings for each.
- m. The default option is to use ‘shall’ in formal requirements, especially for requirements that are defined as ‘must have’ under MuSCoW rankings.

## EVALUATION OF REQUIREMENTS

In order to be useful, requirements do not only need to be identified, represented and documented, but it is also critical to evaluate how well the proposed and final designs have met the requirements. Ideally, the individual documented requirements will include criteria explaining how they can be evaluated as met, but it is also likely to be necessary to undertake some holistic testing of the system, for example the ease and speed of use in completing a scenario (see Hoffman, 2010, for discussion of possible metrics).

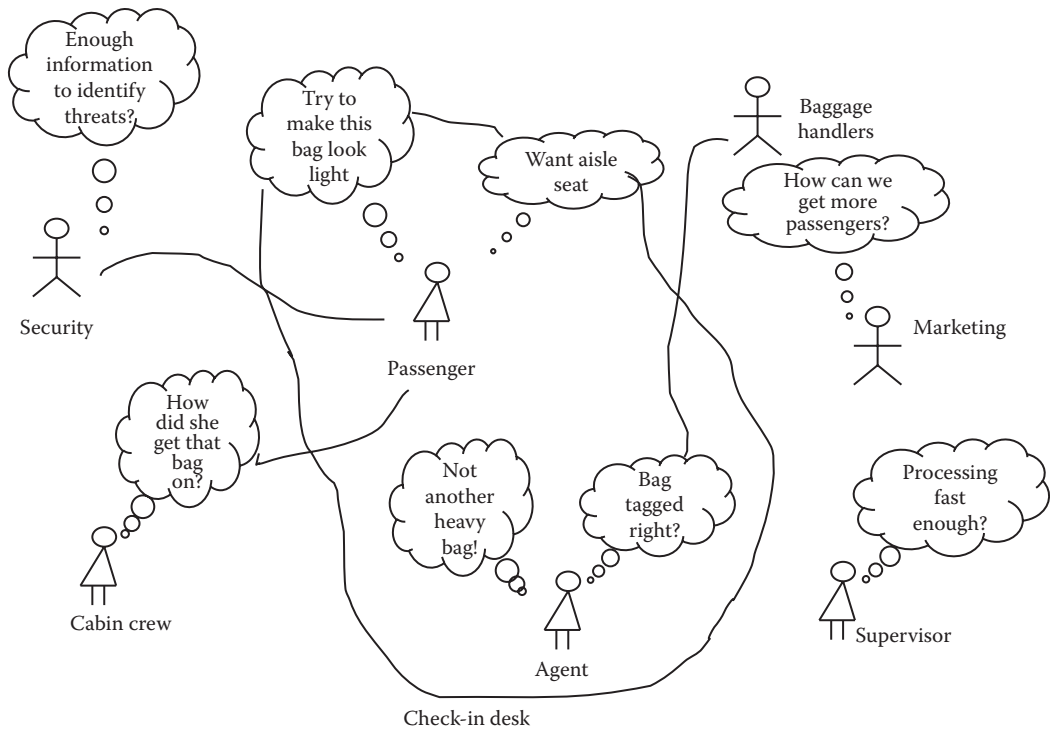
## SYSTEMS ANALYSIS AND REPRESENTATION

### STAKEHOLDER ANALYSIS

A stakeholder analysis identifies everyone who will impact on, or be impacted by, the system. For example, stakeholders in a rail control system might include the following:

- End users of the system (e.g. signalling and control staff)
- End users of the system outputs (e.g. train drivers, passengers)
- Supervisors
- Managers
- Technical support (maintenance, quality assurance, safety, etc.)
- Contractors, suppliers and their staff
- Other industry partners
- The community
- Unintended bystanders

It can be useful to typify stakeholders by various layers of E/HF consideration – for instance, in an ‘onion model’ (Alexander and Maiden, 2004, p. 7) showing the product at the centre and then,



**FIGURE 9.9** A rich picture of needs and wants boarding an airline.

in successive layers, the human–machine system, the socio-technical system and the wider environment (the ‘outside world’). For each of the stakeholders, the ergonomics team will want to know about their goals, relevant functions and roles and responsibilities.

As part of a full stakeholder analysis, the different and sometimes competing needs of different stakeholder groups will have to be reconciled; rich pictures or other pictorial techniques might be used for this, utilising stick figures and thought bubbles to provide instant recognition of any discrepancies and conflicting needs and wants (Checkland, 1981; Checkland and Scholes, 1990; see Figure 9.9). These are discussed in more detail later.

The stakeholder analysis may include definition of personas, usable later in story, scenario and use case generation. Personas are imaginary users, or models or archetypal descriptions of one or more people using the system to achieve specific goals (Pew and Mayor, 2007). They are said to support shared representations of stakeholders across the development team, especially when based on the sort of extensive analysis described in this section.

A very simple example of a persona for passenger activities is shown in the following box:

#### Simplified example of a persona

- Nick Peterborough: Occasional business and leisure traveller.
- Has a company car, prefers to drive than take the train.
- Always has his mobile phone and laptop with him.
- Loves restaurants, watching football, going to the pub.
- Wants include a seat and space to work, refreshments throughout the journey, accurate information throughout the journey and car parking close to the station.

## SYSTEM CONCEPTS AND VIEWS

### Stories, Scenarios and Use Cases

One way to represent how a system might work is to use some sort of rich representation; these may be stories, scenarios or use cases. They are developed to help users visualise possibilities for a new system (or modification of an existing system) and may also be used at a later stage to communicate user requirements to developers in a rich but reasonably systematic fashion. They may also be useful in linking requirements to useful specifications and will help with the early development of evaluation methods. Scenarios, user stories and use cases can all be produced in a variety of forms and to some extent are similar. One main difference is that the notion of scenarios and stories emerged from specialists in interactive system design and product design, explicitly those with an E/HF background, whereas the notion of use cases emerged from the discipline of software engineering and software design. Thus, the former tends more to be descriptions of how a system is or will be used, whereas the latter tends to be more formal descriptions of the functionality needed for a system.

Stories are just that, rich interesting textual descriptions of prototypical people (or personas) undertaking activities in context. In some views, they are a defined type of representing requirements, but for most, they are one type of scenario, probably the most widely used form. That is, stories are seen by some designers as a precursor to scenario production and by others as a type of scenario. Ideally, they describe how people behaved with the old system or product and then hypothetically how they may behave with the new one. They can be written based on what current or potential users say about the reasons for using a system, how they go about it, who they are and what is the context. They frequently take the form of a chatty description of particular uses or circumstances. Stories may be individual (one informant) or composite (constructed across multiple informants), and may be based on current circumstances or anticipate the future.

Scenarios are rich context-based descriptions to visualise possibilities and communicate user requirements to developers in a very accessible fashion. They provide a structured or expanded version of stories, defining person/people (or actors), products (props), tasks (actions/dialogue) and environment/context (staging), and should usually be set within a business system boundary to ensure that they do not grow too extensive. In most cases, scenarios will include normal and abnormal operating conditions, the latter possibly defined and modelled as exceptions or opportunities for failure such as errors and violations (e.g. Maiden et al., 1999; Viller et al., 1999), in which case the failure modes and effects, fault tree and general causal analyses will be valuable. Scenarios may include various user profiles (or personas) and 'what-if' situations.

Stories and scenarios should be regarded very much as a pragmatic means to an end, for exploring context and user needs and then communicating requirements in an accessible form and vocabulary for use by all. Scenarios are often used as a precursor to developing lists of (molecular) requirements, especially in larger system development, but for small pieces of equipment or products, they may be the only type of requirements produced. Scenarios can also aid dialogue and understanding amongst the development team and stakeholders. For instance, developers might construct system potential options, based on scenarios, in video/animation, etc., to present to users to stimulate their creativity and to extend the vision they have of the technology possibilities.

Strongly related to scenarios are use cases. Use cases are usually very structured forms of scenarios, sometimes in a formal language and within a somewhat rigid defined format. Use cases are usually produced in linked series and contain the same or very similar elements to scenarios, with somewhat more connection to business events which trigger information flows and drive the system processes. In simple terms, a use case is a description of a sequence of actions or events for a specific case of some generic task to be accomplished by the system (Maiden et al., 1999). In more detailed versions, a use case template of the content may include functional goal; primary and other roles (actors); main, normal, good day and bad day scenarios; variation scenarios; exceptions; preconditions including the trigger for the start of the use case; guarantees or post-conditions, what must be true when the use case ends; stakeholders and their interests; and local qualities and constraints (non-functional requirements).

## Formal System Representations

In many cases, it might be useful to make more formal representations of the operational view (how the system is or will be used), physical view (what the system actually contains or will contain) as well as functional view (what the system does or will do) – see Chapanis (1996, p. 41). While this could simply use hand-drawn diagrams, there are also options that allow integration with the wider systems engineering process.

Enterprise architectures are increasingly popular methods for modelling large systems that provide a wide-ranging set of interlinked ‘views’ of a system model (operational, technical, strategic, etc.) for the purposes of understanding complexity, planning change and elucidating requirements. Indeed, within the MoD Architectural Framework implementation, there has been significant effort given to providing human views that map domains of E/HF into and onto the larger systems model (see Bruseberg, 2008). In other settings (especially software development), another popular modelling technique, that of UML (Unified Modelling Language) developed by Booch, Rumbaugh and Jacobson (Booch et al., 1998; Fowler, 2003), may also be used to integrate these views with representations of the wider systems engineering and technical elements of the project. UML contains, for example, a standard format for representing use cases.

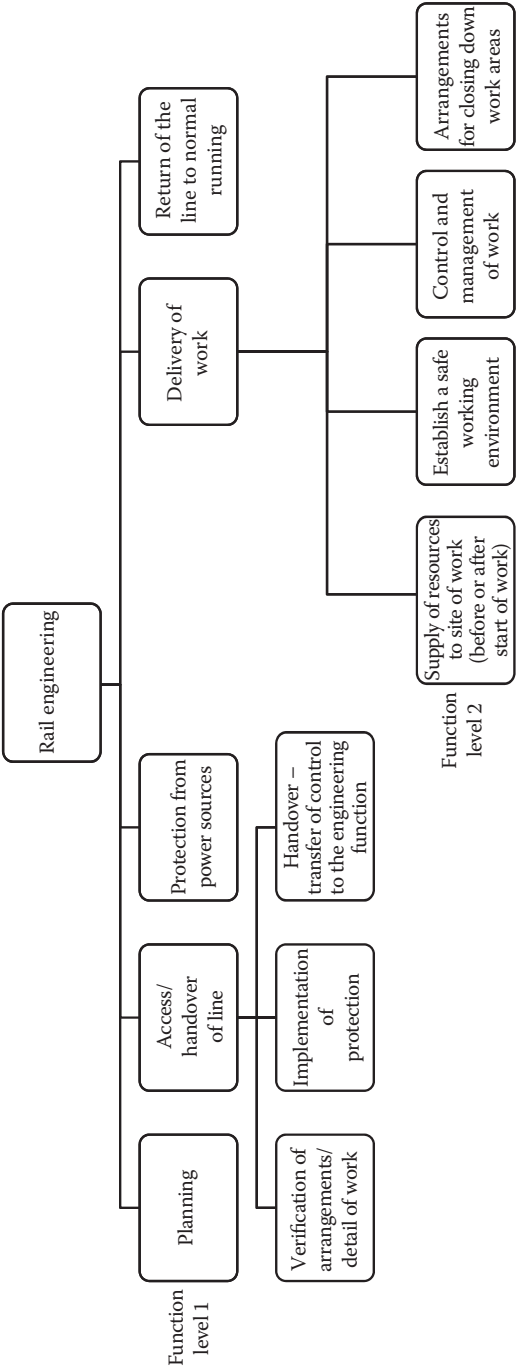
WESTT (workload, error, SA, time and teamwork) is another approach using multiple views and analyses to represent systems, which was developed by the U.K. Human Factors Integration Defence Technology Centre in the mid-2000s in response to the challenge of understanding complex distributed work in emerging military systems (Houghton et al., 2008). The basis for this approach is analytical prototyping within three linked networks; tasks, knowledge and social. Thus, an individual actor can be seen as located within three networks simultaneously: they are involved in tasks linked to other tasks, they possess a certain amount of knowledge linked to other knowledge held by agents and actors in the system and they occupy a position within a social network of communications and collaboration. Thus, the optimisation of the system as a whole can be understood in turn as the co-optimisation of the three networks. In addition, the tool also generates linked visualisations some of which are found in UML (e.g. use cases). Houghton, Baber and Chaudemanche (2008) have also proposed a methodology for using WESTT as a method for integrating data collection and analysis with the generation of UML class diagrams of the wider system.

A final emerging form of systems representation (which can be used at the conceptual stage for communicating and exploring how a system might function or to provide a formal model of a system for detailed analysis) is to use a system dynamics approach. Such models consist of nodes representing system components or qualities and signed arcs that show their interconnection with other parts of the system. Such an approach is argued to capture the interactive complexity of a system and moves the analysis of systems beyond their static structural representation into an appreciation of how the structure will change over time and the dynamics behind those changes (see Leveson, 2011, for a comprehensive review and application in exploring system safety).

## SYSTEMS ANALYSIS

### Function Analysis

A critical contribution to understanding current and potential future work systems is work analysis. One approach to this is a function analysis. Importantly, at this early stage of development, there should have been few, if any, firm decisions made about the form of technology – hardware and software (although one exception is where the design is to integrate a predetermined piece of equipment into an existing setting). A key analysis is of what must be done to achieve the system goals – what functions must be completed – and a function analysis usually describes these and their interactions in a hierarchical, tabular or graphical representation. Figure 9.10 is an example of a first top-level function analysis for rail engineering work, and Table 9.2 shows an example line from a more detailed analysis. Greater detail can be included in graphically represented functional flow diagrams (Figure 9.11).



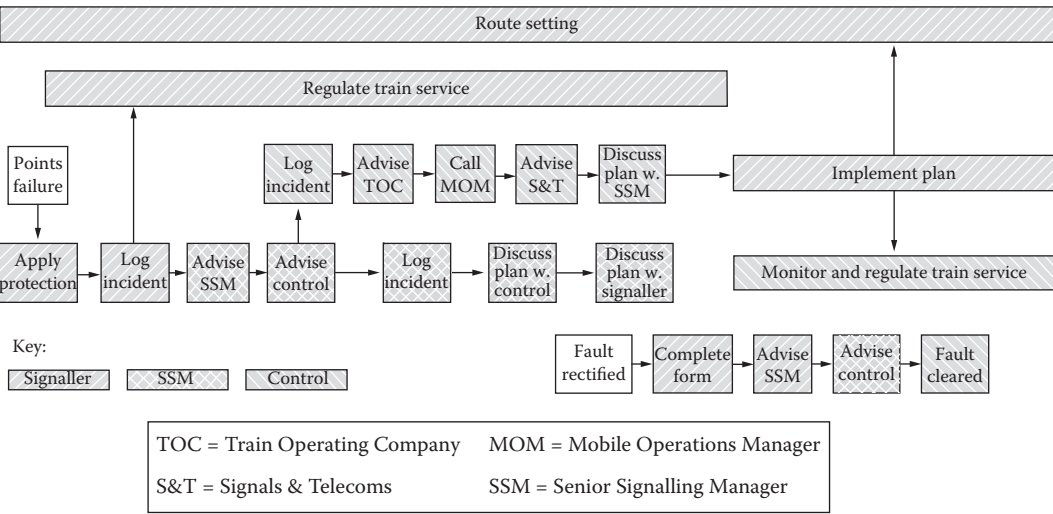
**FIGURE 9.10** Extract from a higher-level rail engineering function analysis. (From Wilson, Ryan, Schock, Ferreira, Smith and Pitsopoulos; 2009.)



**TABLE 9.2**  
**One Line Extract from a Detailed Rail Engineering Function Analysis**

| Function        | Sub-function                    | Activity                              | Sub-activity level 1       | Sub-activity level 2 | Major issues; systems changes | Factors affecting; assumptions   | Study needed   |
|-----------------|---------------------------------|---------------------------------------|----------------------------|----------------------|-------------------------------|----------------------------------|----------------|
| Access/handover | Prepare for transfer of control | Signaller checks possession documents | Signaller check for delays | Check other boxes    | New system more flexible?     | Depends on the number of permits | Comms analysis |

Source: From Wilson, Ryan, Schock, Ferreira, Smith and Pitsopoulos (2009).



**FIGURE 9.11** Sample function flow diagram for route setting. (Adapted from Balfe, N. et al., Applying human factors in the design of future rail systems, in *Human Centred Automation*, de Waard, D., Gerard, N., Onnasch, L., Wiczorek, R., Manzey, D., eds., Shaker Publishing, Maastricht, the Netherlands, 2011, pp. 239–250.)

**Task Analysis**

For some ergonomists, task analysis is the central core of their activity. In the design of some large systems, the task analysis may be the complete specification or blueprint of systems design, especially in highly proceduralised industries. In such circumstances, enormous efforts may go into delivering a very formal, usually hierarchical, task analysis. On the other hand, for smaller (product) design projects, the task analysis may be a much less formal set of representations, covering only the most salient product functions (see Hahn et al., 1995). The methods used to gather information to produce such different types and level of task analysis may be any covered in this book, but will at the least usually involve some form of interviewing and some form of observation.

In its original form, task analysis was principally used to help determine training programme content (e.g. Annett and Duncan 1965). Therefore, one view of task analysis is that it is strictly a representation of what must be done to achieve system goals, no matter how (i.e. current performance is, at most, only marginally relevant). In more recent times, task analysis has also represented how work, or human–machine systems interactions, is performed in practice. Vicente (1998), using parallels with decision-making research, has very usefully distinguished normative work analysis (what must be done) and descriptive work analysis (how work is actually done); he also proposes

formative work analysis as a means to move us from rich accounts of what is going on to valuable contributions in design. More detail on task analysis can be found in Chapter 6.

### Allocation of Function

Since the earliest days of proposals for a systems ergonomics approach to dealing with human factors in systems design, there has been discussion of how and when to make decisions about the functions that people will carry out. From the 1950s onwards, this was known as allocation of function. In earlier years, the approach of allocation of function was inseparable from the idea of Fitts lists (Fitts, 1951): Men are better at...machines are better at (MABA–MABA). Subsequently, human factors specialists in design have seen that the decisions over which people/jobs might be responsible for which functions could also be taken in a systematic fashion in the design process – sometimes called division of function. Then, a third generation of thinking about function allocation came about with the increased use of automation and of more flexible computer systems, to give ideas of flexible or dynamic allocation of function.

The earliest (misunderstood) ideas on task allocation based on comparative ability have been expanded since (e.g. Bailey, 1989) to include as possible guiding philosophies: strict comparability – ‘people are best at/machines are best at’ (PABA–MABA); the search for – or avoidance of – leftover design (i.e. designers leave for the people everything computers cannot do no matter what whole job this might leave); economic allocation (i.e. make allocation decisions on cost considerations only); more positively, allocation and subsequent systems design which provides for rounded skilled jobs; and flexible or dynamic allocation of function.

The text book approach (e.g. Sanders and McCormick, Bailey, Chapanis) suggests that we carry out a detailed task or at least function analysis and then assign each activity to be completed by person A, person B, etc., or machine/computer X, machine/computer Y. In this approach, traceability is provided by recording what criterion was used in order to make the allocation. This would not necessarily be, as in the original Fitts list (Fitts, 1951), whether the person or computer/machine was actually the best to carry out that activity since there has long been acceptance that this leads to a piecemeal type of design. However, this molecular approach of allocating each task in turn is still recommended, albeit using criteria such as cost, maintainability, logistics support, environmental impact as well as performance attributes to do with time and accuracy. Criticisms of Allocation of Function (AoF) and Fitts lists, or at least their rigid and almost automatic application, are well rehearsed. Their military origin means they may not be applicable for civilian systems design where people have more leeway; they are clearer on quantitative attributes providing a bias to the non-human system elements whose advantages tend to time and quantity measures; psychosocial needs are not considered; there is no allowance for learning of the people, thus making early design decisions potentially redundant; and they will lead to ‘out of loop of control’ jobs. The summary of these criticisms is usually that people and computers should be complementary, not competitive (Grote et al., 1995; Jordan, 1963; Majchrzak, 1988), which fits with the ideas of flexible or adaptive allocation. However, even within a flexible system, the tasks still need to be allocated.

Dynamic or flexible AoF suggests that functions are allocated dynamically as goals, tasks or situations require. Support or control by the computer can be ‘on call’, ‘offered’ or automatic, and the function of the computers can be anything from re-forming displays of information, to offering tips, through to decision support and onto full automation and control of all processes (except when it goes wrong!). Computer support within the automation may be used to kick in during situations of high loads on the operators, during emergencies to take over normal running of other elements of the total system, to make complex calculations, to automatically display checklists or procedures displays, to provide multiple alternate display formats, to carry out prediction and to present alternatives in decision making. The difficulty is how to implement this automation without increasing the complexity for the operator working with an already complex system.

In practice, this formal application of allocation of function is very difficult. Fuld (1993) gives three reasons for this difficulty: first, allocation of function is not a distinct or coherent process but is

implicit in the process of design; second, systems are increasingly complex and evolve over time, limiting the value of allocation based on predictability; and third, humans are the most flexible yet also the least standardised component, meaning that design is always uncertain. While it is still important to carry out the function allocation process as formally and in as defensible way as possible, it is often necessary to work more with a process akin to the forming of roles (human and computer) in a holistic fashion. A possible analogy to illustrate this is the difference between user requirements specification via tabular or other lists (allocation of function) and via scenarios (roles development).

Our own experience in supporting development, enhancements and renewals of rail control and infrastructure projects is that in real projects, the human factors team within the development process tends to go for a mixture of a leftover approach, good job design (or at least arguing the case for this) and cost. The reason for the leftover approach is a constraint that we often find that to get a project off the ground quickly, technology decisions are made very early in the life cycle, and we are left playing catch up and trying to pull back and find a better role for people by taking some activities back off the computers through rational argument over the feasibility of unbalanced automation and the difficulties this presents for work during perturbations and emergencies (including recovery). Whatever the doubts and criticisms over formal task allocation and of allocation of function, the human factors team in any systems design project will, at some point, have to answer the following kinds of questions from their colleagues in design, engineering, systems analysis and operations: What should person A or person B, best do? Can the roles of person A and B be combined? When and how should a computer/automation take on current human roles? When and how should a person override the computer (in a supervisory control agenda)? How much can person A actually do, safely, reliably, efficiently, that is what is a reasonable workload? and How much can persons A and B do in collaboration, and is this clearly less or more (and better or worse performed) than A and B separately – that is does a gestalt apply or not?

A broad framework for the allocation of functions in system design would consist of five stages:

- Functional analysis – what must the system achieve?
- Operational analysis – how will the system work?
- Definition of user and use – who will work within the system and how?
- Definition of circumstances of use – where and why will the system operate?
- Assessment of normal, abnormal and emergency activities.

The initial work can be achieved by an integrated design team, guided by a human factors specialist, but an important step is to validate the job designs with a wider group of stakeholders, including operators and maintainers at a minimum. This validation exercise can make use of visualisations of the roles, role-playing, serious games or use cases to engage the stakeholders and generate useful feedback from the session which may modify the allocation of functions.

## FINAL THOUGHTS

Of course, if we just mechanically follow the advice in this chapter, or indeed in this book as a whole, we will not necessarily produce a good systems design. Nor will the guidance even ensure that all human factors are accounted for during development. The ergonomist and the E/HF team will often use the craft of E/HF as much as its science in supporting, sometimes leading, systems development projects. But good E/HF design will require well-informed choices to be made – about functions, the role of people and the requirements and consequences for them, and for these choices to be made in a systematic manner. Carrying out requirements specification, intelligence gathering, function analysis, task analysis and user analysis, and where warranted implementing an HFIP will improve the human factors in systems development and in the design of human-machine systems. However, it should be remembered that analysis and modelling is not *design*. Any analysis method or modelling tool selected must provide evidence that is useful in making design decisions.

Systems analysis and design poses a significant challenge to E/HF practitioners on multiple levels. It requires marshalling a varied range of methods, attempting to comprehend situations of considerable complexity and, as we have discussed at various points in this chapter, may also test interpersonal and managerial skills. While it is relatively straightforward to describe methods and approaches as we have done in this chapter, there remain key elements of systems E/HF that are somewhat harder to offer specific advice about, but that success or failure may well rest upon: it will likely not be enough to be merely a competent practitioner of E/HF but also an integrator, a leader, an explainer, a negotiator and, ultimately, an ambassador for our discipline. As Wilson (2014) notes: ‘The point is that wherever the E/HF function is situated it must collaborate with all other functions in order to do its job. It is often because a good systems-oriented E/HF practitioner actually does work with all other functions, indeed has to do this in order to do their job properly, and must understand stakeholder needs (including those of people managing or building the system as well as end users), that they are invaluable for successful systems engineering’.

## REFERENCES

- Aikin, C., Rollings, M. and Wilson, J.R. (1994). Providing a foundation for ergonomics: Systematic ergonomics in engineering design (SEED). *Proceedings of the 12th Congress of the International Ergonomics Association*, pp. 276–278. Toronto, Ontario, Canada, 5 August 1994.
- Alexander, I. and Maiden, N. (2004). *Scenarios, Stories, Use Cases: Through the Systems Development Life-Cycle*. Chichester, U.K.: John Wiley & Sons.
- Annett, J. and Duncan, K.D. (1967). Task analysis and training design. *Occupational Psychology*, 41, 211–221.
- Bailey, R.W. (1989). *Human Performance Engineering: Using Human Factors/Ergonomics to Achieve Computer System Usability*, 2nd edn. Englewood Cliffs, NJ: Prentice Hall.
- Balfe, N., Lowe, L., Hillier, G. and Nock, P. (2011). Applying human factors in the design of future rail systems. In D. de Waard, N. Gerard, L. Onnasch, R. Wiczorek and D. Manzey (eds.), *Human Centred Automation* (pp. 239–250). Maastricht, the Netherlands: Shaker Publishing.
- Blanchard, B.S. and Fabrycky, W.J. (2011). *Systems Engineering and Analysis*, 5th edn. Upper Saddle River, NJ: Pearson Education.
- Booch, G., Jacobson, I. and Rumbaugh, J. (1998). *Unified Modeling Language User Guide*. Reading, MA: Addison-Wesley.
- Booher, H.R. (2003). *Handbook of Human Systems Integration*. Hoboken, NJ: Wiley.
- Bruseberg, A. (2008). Human view for MODAF as a bridge between human factors integration and systems engineering. *Journal of Cognitive Engineering and Decision Making*, 2(3), 220–248.
- Chapanis, A. (1996). *Human Factors in Systems Engineering*. New York: John Wiley.
- Checkland, P.B. (1981). *Systems Thinking, Systems Practice*. Chichester, U.K.: John Wiley.
- Checkland, P.B. and Scholes, J. (1990). *Soft Systems Methodology in Action*. Chichester, U.K.: John Wiley.
- Cherns, A. (1987). Principles of sociotechnical design revisited. *Human Relations*, 40, 153–162.
- Clegg, C.W. (2000). Sociotechnical principles for systems design. *Applied Ergonomics*, 40, 463–478.
- Federal Aviation Administration. (2003). Guidelines for human factors requirements development, Report AAR-100. Office of the Chief Scientific and Technical Advisor on Human Factors.
- Fitts, P.M. (1951). *Human Engineering for an Effective Air-Navigation and Traffic-Control System*. Columbus, OH: Ohio State University.
- Forrest, E. and Fletcher, G. (2009). Developing the HFI social and organisational domain: Final report. HFIDTC Technical Report HFIDTC/2/WP2.18/3. <https://www.defencehumancapability.com/Portals/0/HFIDTC/HFI%20Process/Phase%202/HFIDTC-2-2-18-3-social-organisational.pdf>. Yeovil, UK: Aerosystems International.
- Fowler, M. (2003). *UML Distilled: A Brief Guide to the Standard Object Modeling Language*, 3rd edn. Reading, MA: Addison-Wesley.
- Fuld, R.B. (January 1993). The fiction of function allocation. *Ergonomics in Design*, 1, 20–24.
- Grote, G., Weik, S., Wäfler, T. and Zölch, M. (1995). Criteria for the complementary allocation of functions in automated work systems and their use in simultaneous engineering projects. *International Journal of Industrial Ergonomics*, 16, 367–382.
- Hahn, H.A., Houghton, F.K. and Youngblood, A. (1995). Job-task analysis: Which way? Three methods for making human resources decisions. *Ergonomics in Design*, 3(4), 22–28.

- Hoffman, R.R. (2010). Some challenges for macrocognitive measurement. In E.S. Patterson and J.E. Miller (eds.), *Macrocognition Metrics and Scenarios*. Farnham, U.K.: Ashgate.
- Hollnagel, E. (ed.) (2003). *Handbook of Cognitive Task Design*. Mahwah, NJ: Lawrence Erlbaum.
- Houghton, R.J., Baber, C. and Chaudemache, E. (2008). Integrating human factors with systems engineering: Using WESTT to design a novel user interface for accident command systems. *Human Factors Society Annual Meeting Proceedings*, 52(4), 1944–1948.
- Houghton, R.J., Baber, C., Cowton, M., Walker, G.H. and Stanton, N.A. (2008). WESTT (workload, error, situational awareness, time and teamwork): An analytical prototyping system for command and control. *Cognition, Technology and Work*, 10(3), 199–207.
- Huang, G.Q. (1996). *Design for X: Concurrent Engineering Imperatives*. Cornwall, U.K.: Chapman & Hall.
- Jordan, N. (1963). Allocation of functions between man and machines in automated systems. *Journal of Applied Psychology*, 47, 161–165.
- Kirwan, B. and Ainsworth, L. (1992). *A Guide to Task Analysis*. London, U.K.: Taylor & Francis Group.
- Kragt, H. (1992). *Enhancing Industrial Performance*. London, U.K.: Taylor & Francis Group.
- Lepreux, S., Abed, M. and Kolski, C. (2003). A human-centred methodology applied to design support systems design and evaluation in a railway network context. *Cognition, Technology and Work*, 5, 248–271.
- Leveson, N. (2011). *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, MA: MIT Press.
- Lindgaard, G., Dillon, R.F., Trbovich, P., White, R., Fernandes, G., Lundahl, S. and Pinnamaneni, A. (2006). User needs analysis and requirements engineering: Theory and practice. *Interacting with Computers*, 18(1), 47–70.
- Luzeaux, D. and Rualt, J.-R. (2008). *Systems of Systems*. London, UK: Wiley-ISTE.
- Maiden, N., Minocha, S., Sutcliffe, A., Manuel, D. and Ryan, M. (1999). A co-operative scenario based approach to acquisition and validation of system requirements: How exceptions can help! *Interacting with Computers*, 11(6), 645–664.
- Majchrzak, A. (1988). *The Human Side of Factory Automation: Managerial and Human Resource Strategies for Making Automation Succeed*. San Francisco, CA: Jossey-Bass.
- Meister, D. (1985). *Behavioural Analysis and Measurement Methods*. New York: John Wiley.
- Meister, D. and Enderwick, T.P. (2002). *Human Factors in System Design, Development and Testing*. London, U.K.: Lawrence Erlbaum.
- Moray, N. (1994). “De Maximum non Curat Lex” or How context reduces science to art in the practice of human factors. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 526–530). Nashville, Tennessee.
- National Audit Office (2010). The failure of the FIRE control project. Report by the Comptroller and Auditor General, HC 1272. London, U.K.: HMSO.
- O’Brien, T.G. and Charlton, S.G. (1996). *Handbook of Human Factors Testing and Evaluation*. Mahwah, NJ: Lawrence Erlbaum.
- Pew, R.W. and Mavor, A.S. (2007). *Human-System Integration in the System Development Process: A New Look*. Washington, DC: Academies Press.
- Ramani, K., Ramanujan, D., Bernstein, W.Z., Zhao, F., Handwerker, C., Choi, J.-K., Harrison, K. and Thurston, D. (2010). Integrated sustainable life cycle design: A review. *Journal of Mechanical Design*, 132(9), 1–15.
- Robertson, S. and Robertson, J. (2006). *Mastering the Requirements Process*, 2nd edn. Boston, MA: Addison-Wesley.
- Sanders, M.S. and McCormick, E.J. (1987). *Human Factors in Engineering and Design*, 6th edn. New York: McGraw-Hill.
- Salmon, P., Stanton, N.A., Walker, G. and Green, D. (2010). Cognitive task analysis – a review. In W.S. Marras and W. Karwowski (eds.), *Fundamentals and Assessment Tools for Occupational Ergonomics*. London, U.K.: Taylor & Francis Group.
- Singleton, W.T. (1974). *Man-Machine Systems*. Harmondsworth, U.K.: Penguin.
- Stasinopoulos, P., Smith, M.H., Hargroves, K.C. and Desha, C. (2009). *Whole System Design: An Integrated Approach to Sustainable Engineering*. London, U.K.: Earthscan.
- Vicente, K. (1998). *Cognitive Work Analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Viller, S., Bowers, J. and Rodden, T. (1999). Human factors in requirements engineering: A survey of human sciences literature relevant to the improvement of dependable systems development processes. *Interacting with Computers*, 11(6), 665–698.
- Waterson, P.E. and Kolose, S. (2010). Exploring the social and organisational aspects of human factors integration: A framework and case study. *Journal of Safety Science*, 48(4), 482–490.

- Wickens, C.D., Gordon, S.E. and Liu, Y. (1998). *An Introduction to Human Factors Engineering*. New York: Longman.
- Widdowson, A. and Carr, D. (2002). *Human Factors Integration: Implementation in the Onshore and Offshore Industries*. Norwich, U.K.: HMSO.
- Wilson, J.R. (2014). Fundamentals of systems ergonomics/human factors. *Applied Ergonomics*, 45(1), 5–13.
- Wilson, J.R., Ryan, B., Schock, A., Ferreira, P., Smith, S. and Pitsopoulos, J. (2009). Understanding safety and production risks in rail engineering and protection. *Ergonomics*, 52(7), 774–790.
- Wilson, J.R. and Whittington, C.M. (2001). Implementation of self-managed teams in manufacturing: More of a marathon than a sprint. *AI & Society*, 15(1–2), 58–81.



---

# 10 Involving People in Design Research

*David Kirk, Ian McClelland and Jane Fulton Suri*

## CONTENTS

|                                                                                    |     |
|------------------------------------------------------------------------------------|-----|
| Introduction.....                                                                  | 252 |
| Purpose of the Chapter.....                                                        | 252 |
| Our Perspective.....                                                               | 253 |
| On Design, Designing and Designer.....                                             | 253 |
| Design as Noun: The Thing Designed.....                                            | 253 |
| Design as Verb: The Activity of Designing.....                                     | 253 |
| Designer: A Person Making Design Decisions .....                                   | 253 |
| Design Researcher.....                                                             | 253 |
| Skills and Attitudes Required.....                                                 | 253 |
| Technical Skills .....                                                             | 253 |
| Social Skills with Participants .....                                              | 254 |
| Social Skills in Design Teams .....                                                | 254 |
| Participants.....                                                                  | 254 |
| Introducing Human-Centred Design.....                                              | 256 |
| Design and Development Process.....                                                | 256 |
| Historical Origins of HCD .....                                                    | 256 |
| Benefits of HCD.....                                                               | 257 |
| Phases of the HCD Process.....                                                     | 257 |
| Explore and Understand Users' Needs, Opportunities and Contexts of Use.....        | 258 |
| Identify Design Options and Directions.....                                        | 259 |
| Create and Refine Specific Solutions and Prototypes .....                          | 260 |
| Evaluate Implemented Designs in the Real World and against User Requirements ..... | 261 |
| Setting Up an Investigation: General Considerations .....                          | 262 |
| Ethical Issues.....                                                                | 262 |
| Objectives .....                                                                   | 262 |
| Participant Goals, Tasks and Activities .....                                      | 263 |
| Location.....                                                                      | 263 |
| Participant Selection.....                                                         | 263 |
| Participant Numbers .....                                                          | 264 |
| Representing the Design.....                                                       | 265 |
| Measures, Observations and Subjective Data.....                                    | 266 |
| Gathering Behavioural Data.....                                                    | 267 |
| Interpretation of Results .....                                                    | 267 |
| Design Criteria .....                                                              | 268 |
| Methods: Overview of Areas.....                                                    | 268 |
| Contextual Observations .....                                                      | 270 |
| Overview .....                                                                     | 270 |
| Behaviour Sampling.....                                                            | 271 |



|                                          |     |
|------------------------------------------|-----|
| Shadowing.....                           | 271 |
| When to Use .....                        | 271 |
| Participant Involvement.....             | 272 |
| Designer and Developer Involvement ..... | 272 |
| How to Use.....                          | 272 |
| Process Involved.....                    | 272 |
| Expertise Needed.....                    | 272 |
| Outcomes.....                            | 272 |
| Examples of Use .....                    | 273 |
| Diary-Keeping.....                       | 273 |
| Overview .....                           | 273 |
| When to Use .....                        | 273 |
| Participant Involvement.....             | 274 |
| Designer and Developer Involvement ..... | 274 |
| How to Use.....                          | 274 |
| Process Involved.....                    | 274 |
| Expertise Needed.....                    | 274 |
| Outcomes.....                            | 274 |
| Examples of Use .....                    | 275 |
| Framework Development.....               | 275 |
| Overview .....                           | 275 |
| When to Use .....                        | 275 |
| Participant Involvement.....             | 276 |
| Designer and Developer Involvement ..... | 276 |
| How to Use.....                          | 276 |
| Process Involved.....                    | 276 |
| Expertise Needed.....                    | 276 |
| Outcomes.....                            | 277 |
| Interviews.....                          | 277 |
| Overview .....                           | 277 |
| When to Use .....                        | 277 |
| Participant Involvement.....             | 277 |
| Interview Formats .....                  | 277 |
| Designer and Developer Involvement ..... | 278 |
| How to Use.....                          | 278 |
| Process Involved.....                    | 278 |
| Expertise Needed.....                    | 278 |
| Outcomes.....                            | 278 |
| Examples of Use .....                    | 278 |
| Questionnaires.....                      | 279 |
| Overview .....                           | 279 |
| When to Use .....                        | 279 |
| Participant Involvement.....             | 279 |
| Designer and Developer Involvement ..... | 279 |
| How to Use.....                          | 279 |
| Process Involved.....                    | 279 |
| Expertise Needed.....                    | 280 |
| Outcomes.....                            | 280 |

|                                          |     |
|------------------------------------------|-----|
| Projective Techniques .....              | 280 |
| Overview .....                           | 280 |
| When to Use .....                        | 280 |
| Participant Involvement.....             | 280 |
| Designer and Developer Involvement ..... | 281 |
| How to Use.....                          | 281 |
| Process Involved.....                    | 281 |
| Expertise Needed.....                    | 281 |
| Outcomes.....                            | 281 |
| Examples of Use .....                    | 282 |
| Role-Playing .....                       | 282 |
| Overview .....                           | 282 |
| When to Use .....                        | 282 |
| Participant Involvement.....             | 282 |
| Designer and Developer Involvement ..... | 282 |
| How to Use.....                          | 283 |
| Process Involved.....                    | 283 |
| Expertise Needed.....                    | 283 |
| Outcomes.....                            | 283 |
| Examples of Use .....                    | 284 |
| Scenario Building.....                   | 284 |
| Overview .....                           | 284 |
| When to Use .....                        | 284 |
| Participant Involvement.....             | 284 |
| Designer and Developer Involvement ..... | 285 |
| How to Use.....                          | 285 |
| Process Involved.....                    | 285 |
| Expertise Needed.....                    | 285 |
| Outcomes.....                            | 285 |
| Examples of Use .....                    | 286 |
| User Trials.....                         | 286 |
| Overview .....                           | 286 |
| When to Use .....                        | 286 |
| Participant Involvement.....             | 287 |
| Designer and Developer Involvement ..... | 287 |
| How to Use.....                          | 287 |
| Process Involved.....                    | 287 |
| Expertise Needed.....                    | 287 |
| Outcomes.....                            | 288 |
| Examples of Use .....                    | 288 |
| User Workshops .....                     | 288 |
| Overview .....                           | 288 |
| When to Use .....                        | 288 |
| Participant Involvement.....             | 288 |
| Designer and Developer Involvement ..... | 289 |
| How to Use.....                          | 289 |
| Process Involved.....                    | 289 |
| Expertise Needed.....                    | 289 |

|                                                     |     |
|-----------------------------------------------------|-----|
| Outcomes.....                                       | 289 |
| Examples of Use .....                               | 289 |
| Future Directions .....                             | 289 |
| Organisational and Technical Barriers .....         | 290 |
| Design as a Strategic Business Advantage.....       | 290 |
| Proving the Value and Relevance of Methods.....     | 290 |
| Earlier Involvement and Faster Response Times ..... | 291 |
| Ecological Validity .....                           | 291 |
| Design Criteria .....                               | 292 |
| Aesthetics and Preferences.....                     | 292 |
| Personal and Sociocultural.....                     | 292 |
| Sustainability Frontier.....                        | 293 |
| Creative and Integrated Methods .....               | 293 |
| References.....                                     | 294 |

## INTRODUCTION

### PURPOSE OF THE CHAPTER

This chapter is primarily about supporting design activity through human-centred research *for* design. It is about how the people we design for can participate in, and otherwise contribute to, the design of things that are useful, usable and enjoyable to use. It is about methods that enable people's needs, desires and values to be effectively embodied and integrated into design activities so that they are embraced by design and development teams.

This book is aimed primarily at practising ergonomists (or human factors engineers). Ergonomics/human factors (E/HF) specialists certainly have the skill base to make use of the methods covered in this chapter, but there are also many other professional practitioners who become the *user advocate* in design teams and who are equally well equipped to use them. Indeed, many of the methods originated outside of ergonomics and are practised by professionals who would never describe themselves as ergonomists. This reflects the reality that the practice of creating satisfactory designs from the end-user perspective requires information and skills that go beyond those that ergonomists have traditionally offered. Specifically, the practice goes beyond concern for a design's functional suitability and usability – its physical and cognitive *fit* – to include consideration of emotional, social, experiential and cultural dimensions (see, for example, discussion in McCarthy and Wright, 2004).

Our focus herein is on the design of devices, services and systems that are intended for use by large numbers of people in the course of their everyday lives, at home, work or play (reflecting the broad applicability of ergonomics work). Consumer products, IT systems, communication tools, living and working environments, vehicles and public transit systems all fall within the scope of this chapter. We will focus on the kind of questions that crop up at the start of a development programme with particular emphasis on innovation rather than simple facelifts for existing products. By *innovation* we mean the introduction of new-to-the-world products and services or radical new design within an existing category. Where an organisation wants to create a *ground-breaking* design, there are greater challenges and opportunities for user-orientated investigations – hence our focus. Of course, even a request to *face lift* a mature product can be significantly informed by the inclusion of insightful data about how a design will be used, in what context and by which users. This is the strength of human-centred design (HCD) research and the methods outlined herein, which will help ergonomists in design teams to support the creative process and to develop better products.

The chapter continues next with our perspective on some key concepts for the reader. This includes outlining the way we use some critical terms such as *design*, *designing* and *designers*;

provides notes on the *design researcher* and the skills they require; and discusses *participants* and their role in the design process. This is followed by an overview of the HCD process as a whole and then the main section on methods at the disposal of the design researcher. We conclude with some observations on directions for the future.

## OUR PERSPECTIVE

### ON DESIGN, DESIGNING AND DESIGNER

We first want to clarify these terms and how we will use them throughout this chapter.

#### Design as Noun: The Thing Designed

We will refer to the *design* as being the *thing* that will be used by people in some context to achieve some particular end. The design might in practice be a product such as a toaster, or a service such as purchasing railway tickets, delivered through some digital technology such as a mobile phone-based app, or it might be an entire system (with multiple stakeholders) such as airline check-in.

#### Design as Verb: The Activity of Designing

The professional practice of designing is increasingly recognised as a multidisciplinary activity and typically involves people of different professional orientations working in design teams of one sort or another (see, for example, discussions by Bannon, 1991 and Rogers et al., 2011). Organisational parameters will determine how, in practice, design teams are set up. Design teams usually involve people working in close collaboration with several other professionals, some of whom will often work together in a formal team and/or sometimes collaborate in more ad hoc and informal ways.

#### Designer: A Person Making Design Decisions

A person working in a professional context whose decisions directly influence the way a thing is designed. This person may, in a professional context, describe themselves explicitly as a designer, or they may be one of a number of several complementary professional communities such as software engineer, interaction designer, user experience (UX) designer/consultant/architect, human factors engineer, usability engineer, applied psychologist and information architect, amongst others. In this context, we often regard ergonomists as designers in that they directly use their expertise to make decisions that contribute to the creation of designs.

#### Design Researcher

We use the term *design researcher* (sometimes framed as a user advocate) to refer to the individual (or people) on a design team who have primary responsibility to ensure that a design is usable, useful and enjoyable for those who will ultimately encounter it. They will take the lead in employing the methods we describe, act as facilitator between a design team and the user community, and be the primary interpreter of the information investigations yield.

## SKILLS AND ATTITUDES REQUIRED

The skills and attitudes needed by the design researcher fall into three categories.

### Technical Skills

All the methods we describe will require, to varying degrees, technical skills in conducting design research studies effectively. These include such skills as formal planning of studies, including coordinating logistics, observation of users and environments, use of various data gathering tools such

as audio and video recording, web analytic software and behaviour logging tools such as wireless sensors, interview and questionnaire design and administration, data analysis (both quantitative and qualitative) and interpretation.

### Social Skills with Participants

Success in drawing useful insights and involving potential end users in the research and design process is not simply a matter of technical competence. The detail, enthusiasm and creativity that people are willing to contribute are dependent in large measure on a design researcher's ability to establish mutual trust and integrity. Generally, there is only a very short space of time in which to develop a positive relationship and to engender cooperation. A good design researcher needs to be able to rapidly engage people. Participants need to understand what to expect and how they will contribute. The design researcher needs to convey a genuine interest in the personal opinions, descriptions and ideas of participants. It is helpful to develop briefing statements and courteous opening remarks and warm-up topics/activities to relax participants and put them at their ease. But nothing beats genuine curiosity, interest and enjoyment of people in these kinds of encounters. Many of the methods invite the design team to use their empathetic skills to identify with potential users' experience of designs, rather than to remain as objective observers and design researchers.

### Social Skills in Design Teams

The design team must embrace the discoveries and integrate them effectively into the developing design. To be effective as a *user advocate*, it is important to be a team player. There are many ways to contribute to a particular team. Some teams are best persuaded by delivering objective factual information and clear specific direction. But other teams respond best when they are personally involved in discovering and empathising with user issues. Which is the best way must be considered case by case, though there are some basics that generally apply. These include an ability and willingness to

- 'See another's point of view' (be empathetic)
- Involve others in your work process
- Understand the technical, organisational and political constraints the team must confront
- Go beyond your own specific assignment, identify with the team's objectives, work collaboratively on the synthesis of results, presentations, report writing and project planning

Several of the methods in this chapter have the effect of *co-discovering* and communicating a shared vision of user issues within the design team. In many cases, the design researcher can take on the role of facilitator and use such methods to help focus the efforts of the design team and make the human issues visible, tangible, believable, manageable and measurable.

## PARTICIPANTS

In the past, there has been some tendency for investigations carried out by E/HF practitioners (and comparable professions) to be modelled to a large extent on formal *experiments* that recruit people who were described as *subjects*. *Subjects* are typically introduced into a context where they carry out some task that is used as a means of *measuring* their response. Referring to people as *subjects* can depersonalise the individual implies that they are just sources of data and that they contribute in rather narrow and specific ways to an *experiment*. In some contexts, designing *experiments* using such parameters may still be appropriate. But in helping us to understand and accommodate what matters to people in the context of design, this type of perspective is often overly limiting.

In this chapter (and in most chapters in this volume), we use the term *participant* to denote a more holistic, and realistic, perspective on the contribution that people can make to design creation. The term participant conveys the notion that people are more active, have agency in the

design process and can collaborate with design teams in positive and creative ways to identify what makes for an appropriate design. In recent years, there has been a growing advocacy of considering people as partners in the design process (co-designing). The strongest movement has come from the advocates of participatory design (which originated with considerations of co-designing working practices and spaces) and participatory ergonomics (see, for example Chapter 34 in this book and Greenbaum and Kyng, 1991). In recent years, there has been a growth of interest in developing these principles to deal with a broad scope of artefacts and user communities; see, for example, Gaver et al. (2004), Druin (1999, 2002), Poggenpohl (2002), Sanders (2002), Suchman (1987) and (from a critical perspective) Greenberg and Buxton (2008).

In addition to their basic capacities and limitations as human beings, participants bring their own experience, expertise and expectations about the products, services and systems they use in their everyday lives. And to an extent, people often design for themselves, a feature of our lives that is often overlooked – see Moran (2002) for a stimulating commentary on people as designers. This chapter is about the methods that can be used to tap into the explicit and tacit knowledge, and feelings that people have, and facilitate the application of their creative energy to the design issue in question.

Despite the many successes of the HCD process, we acknowledge that there are still many developer organisations that are sceptical about the value of including people in the design process. Common negative prejudices are reflected in statements like ‘people lack the appropriate insights’, ‘they lack imagination and cannot be expected to “envision the future”’, ‘people cannot tell you what they want’ and so on. It is not easy for people to imagine the future in the abstract, and because of a lack of know-how and expertise, people often find it difficult to be precise about what they would like or expect. But on a daily basis, people expend a great deal of energy and imagination designing aspects of the world they inhabit, whether it be the way they dress or the configuration of environments in which they live and work. And they continue to adapt and change things as their lives and needs evolve, often exhibiting remarkable levels of imagination in the process.

Experience shows, time and again, that skilled design researchers are able to harness this creativity and provide opportunities for people to participate in professional design activities in productive ways. The goal of the design researcher is to create *dialogue* between end users of a system, broader stakeholders and the design team and, through this dialogue, to illuminate the design issue in question. The dialogue should enable the knowledge, creativity and energy of the various participants to be applied in a focussed and productive way, bringing together and combining the tacit knowledge of system users with the broader expertise of other stakeholders. Design researchers can, using the type of methods described in this chapter, create circumstances which provoke the imagination of participants and help them to reflect upon their behaviour, experience and values, to describe their product and service expectations, and to articulate what is likely to please them, interest them and what they might buy. Tapping into this kind of information allows design researchers to identify and interpret the critical human- and context-focussed knowledge that needs to be taken into account in design projects. Dunne and Raby (2001), Gaver et al. (2004) and Hutchinson et al. (2003) provide examples of how artefacts can be used to provoke the imagination and creativity of participants.

Finally, participant involvement is not a *panacea*. The involvement of participants does not in itself guarantee that designs will be beyond criticism and without faults (see Greenberg and Buxton, 2008 for further discussion). Nor does it guarantee startling breakthroughs. Significant innovations in design come from many sources, and so the methods described in this chapter will benefit from other complementary forms of enquiry. But participant involvement can provide a critical and valuable stimulus. Participants can help make the difference providing that (1) their attention is focussed by the design researcher on the questions participants can answer and on activities that reveal useful insights and (2) the design researchers (and the design team) use their professional expertise to carefully analyse, understand and interpret the information that participants provide. Participants form a critical reference for the design team because they are external to the team and the developer organisation, and because their point of view is fundamentally usage-centric.

## INTRODUCING HUMAN-CENTRED DESIGN

In this section, we describe HCD in the context of the design and development process, discuss a number of general issues concerning setting up an investigation (a design research project) and finally overview the methods.

### DESIGN AND DEVELOPMENT PROCESS

All organisations that develop designs have some form of *process model* that enables them to coordinate efforts and achieve a desired result in a managed way. Typically, the processes adopted by industrial and commercial organisations are determined by various considerations such as commercial objectives, technical requirements and resource management policies; put simply, every company has their own *in-house* way of doing things. For example, in many industries and at many companies, there is now a strong push towards using *Agile* software development methods for IT projects. This approach dictates a very specific way of working, using rapid iterative loops of development produced by small working teams. Many design researchers who wish to work with an HCD approach find that they need to integrate their practices and adapt their methods to suit the *Agile* approach (see Bowles and Box, 2010 for specific discussion on combining user experience work with *Agile* processes).

The goal of HCD is to focus attention on the critical human issues throughout the design and development process so that the inevitable trade-offs between human, commercial and technical issues can be managed in a balanced way. If the human issues are not articulated in a clear and manageable form, they are likely to get very easily pushed aside when the pressure of costs, schedules and technical issues come to bear. To be effective, the activities within an HCD process need to be aligned with the *host process* (such as the *Agile* development environment mentioned earlier). The key to success is to ‘ask the right questions at the right time’ and ‘deliver the right information to the right people when they need it’. Ideally, HCD activities will be fully integrated into an organisation’s *process model*, and a key part of any HCD project is to start by uncovering this model. It is advisable to understand the receiving organisation’s process and at what points human-centred contributions will best complement and influence outcomes. In any case, knowledge of the host process in general, and the specific phase(s) where a design researcher plans to contribute, often has a big influence over the choice of method(s) for a particular investigation.

HCD is *not* just fantasising about usage. HCD practitioners need to *gather* evidence of what is appropriate in usage-centric terms, to *interpret* this evidence in terms of the decisions that confront the design team and to *communicate* the evidence in a form that the design team can use. See, for example, Curtis et al. (1999), who report on a project aimed at communicating customer-focussed design data within a large industrial organisation.

### HISTORICAL ORIGINS OF HCD

HCD has emerged over the last 50 years as a fundamental component of E/HF work. The methods that we used to understand technology users are continually evolving as evidenced by numerous recent books describing extant methodologies (Hanington and Martin, 2012; Lazar et al., 2009; Rogers et al., 2011) and in a variety of academic papers proposing and evaluating new user research techniques (see Vermeeren et al., 2010 for a good overview of academic work). Much of this development in technique has stemmed from HCD practitioners’ recognition of the important contribution that other academic and professional disciplines (such as sociology and anthropology) can make to the design process. In particular, a closer engagement with those trained in the creative arts and design has led to increased adoption of more empathetic user-research approaches (see, e.g., Isbister et al., 2006).

Within the specific field of human-computer interaction (HCI), the development of techniques and orientations to users has been described as a series of *waves* (akin to Kuhnian paradigm shifts).



The first wave of HCI was characterised by traditional human factors research, taking experimental methods and traditional approaches to *usability engineering* (Faulkner, 2000) as the basis of user engagement. The second wave, sparked by Bannon's (1991) seminal work *Human Actors not Human Factors*, ushered in an era of greater consideration for the sociological underpinning of technology use and the collective and situated (context specific) nature of technology interaction (this incorporated more participatory approaches to design), a perspective which had previously been given scant regard in system development. The third wave of HCI, however, as signalled by authors such as Susanne Bodker (2006), has sought to develop techniques for user research which probe more value-laden user concerns, attempting to understand, for example, how a design might meet a user's requirements for qualities such as aesthetics, engagement, fun and emotionality (see Blythe et al., 2003 on *Funology* and Cockton, 2004, for a further discussion of value-centred approaches to HCD).

In recent years, there has been an associated development of an International Standard (ISO 9241 Part 210: 2010 – 'Human-centred design for interactive systems'). This standard provides a process through which HCD techniques can be deployed (and is elaborated upon further in the rest of this chapter). In part, the HCD process has been defined to help position HCD activities within prevalent models of system and software development otherwise used by technology developers (see Sommerville, 2010).

## **BENEFITS OF HCD**

In broad terms, it helps to

- Inspire new ideas and design directions
- Create new paradigms and value for existing product or service offerings
- Increase real and perceived value of a design
- Provide better experiences for people using a design
- Avoid wasting development effort on a bad design idea
- Make a weak idea much better and turns a good idea into a great one
- Reduce exposure to product liability issues
- Reduce customer complaints and product recalls
- Create a more efficient design process in relation to
  - Faster and more precise definition of functionality
  - Faster choice of appropriate interaction technologies
  - Earlier determination of design performance targets
  - More cost-effective way of assigning design team effort

For those who wish to read more about HCD and its development and methods as applied to interactive system design, alongside arguments for its strategic importance to business and innovation, there are several authors worth investigating (see Buxton, 2007; Kolko, 2011; Moggridge, 2007; Norman 2010; Shneiderman, 2003).

## **PHASES OF THE HCD PROCESS**

There are many ways of describing the HCD process. For the sake of simplicity, we have chosen to present it as four key phases in evolving a design (based in part on ISO 9241 part 210):

1. Explore and understand users' needs, opportunities and contexts of use
2. Identify design options and directions
3. Create and refine specific solutions and prototypes
4. Evaluate designs in the real world and against user requirements

In reality, the HCD process is iterative, and the phases are not strictly sequential and self-contained as this description implies. But this will serve as a useful structure to explore the kinds of questions that a design team may have to deal with as they make progress through a design programme.

## Explore and Understand Users' Needs, Opportunities and Contexts of Use

### *Objectives*

The purpose of this phase is to discover what matters to people that could be better supported through design, and what might motivate or inhibit the appeal and use of a future design. The questions, implicitly or explicitly, may be

- What are the key characteristics of expected users and contexts?
- What are their unmet needs in these contexts?
- What are the appealing aspects of an activity or existing design? What are the frustrations?
- What would be desirable characteristics of a design that addressed these aspects?
- What are people's expressed and latent needs? What is their relative importance?

*Example: Your team is asked to come up with design concepts for a new in-car information system.*

- What different kinds of information might people need or want in a vehicle?
- How do the needs of a busy parent, a travelling salesperson or a holidaymaker differ?
- What are the delights and frustrations of people's current experience? How could existing methods or tools be improved?
- What are the characteristics of a product/service that would enhance people's experience?

Although the design team probably already has some type of design in mind, the focus of an initial investigation should not be directed upon this specifically. Rather, the focus should be on understanding which group of people are expected to use the design, the context in which it will be used, key motivators that create the perceived need and the criteria that people will likely use to judge the value of the design.

### *How People Can Contribute*

By examining their current experiences, people can almost always provide valuable insights into benefits they will seek in future designs. Exploring their current experience can also reveal why certain designs are, in one-way or another, inappropriate. A useful approach is to explore, together with their users, weaknesses or inadequacies of current designs, and identify what remedies would be appropriate. Such exercises act as catalysts in revealing underlying motivators and sought-after benefits. People can also contribute to an understanding of the attributes of a design or the usage context that will ensure user satisfaction.

Understanding of potential users, their needs and the use context enables the design team to begin to identify a hierarchy of needs and characterise key design challenges and opportunities as well as establish initial design criteria. To do all these, the design team needs to interpret the information in terms of user requirements:

- Those requirements that the design aims to meet
- Those requirements that will not be met
- Those requirements that may be better supported by some complementary device or activity

The outcomes of this phase should enable the design team to profile the people who are expected to use the design and the circumstances under which the design will be used. This should also lead to the generation of preliminary usage scenarios based on the evidence that the design team has been able to gather. The usage scenarios encapsulate descriptions of typical users, typical usage situations, and the range of tasks that users will be expected to undertake. The usage scenarios might be textual descriptions (narratives of use), or quite commonly, they might be graphically represented, using Use Case Diagrams (Sommerville, 2010), Storyboards (Buxton, 2007) or Rich Pictures (Monk and Howard, 1998). In short, usage scenarios characterise how the design team expects the design to be used in practice. Usage scenarios are the first steps in generating particular solutions.

## Identify Design Options and Directions

### *Objectives*

The goal of this phase is to respond to the discovered needs and opportunities by generating and representing design options. And from these options, select directions for further development.

Questions that are commonly addressed in phase 2 will include the following:

- What new design ideas emerge from understanding people's needs and desires?
- How do the proposed ideas map onto the anticipated usage situations?
- What functional and technical attributes could enhance usage?
- How will the design fit within contextual constraints: physical, sociocultural and socio-technical?
- What attributes should the design have to meet the critical acceptability factors?

*Example: Your team is tasked with exploring ideas for a new health-drink-making machine for use in cafes and restaurants.*

- How might the configuration of a machine enhance the operator's workflow and interactions with colleagues and customers?
- How might the machine be positioned in relation to other equipment?
- Is it more appropriate to conceal the operation of the machine or to feature it?
- How might this influence machine form-factor to be as efficient, safe and comfortable as possible in use?

In practice, there is substantial interplay between phase 1 and phase 2. Often, the design team already has a good idea about possible design directions before setting out to explore needs. Nevertheless, it is also usually the case that, whatever the early ideas, the design team will want to explore other possibilities in response to the new evidence available to them.

### *How People Can Contribute*

People can easily highlight features of a design that are attractive as well as those that are not. People are often able to identify the characteristics that are desirable and those to be avoided. But to achieve this kind of contribution, design concepts or new technologies need to be *experienced* by people in tangible ways that clearly relate to their own context. Given this, participants can provide relevant detail and texture about the usage context and help identify the strengths and weaknesses of particular design proposals. Combined with the skill and knowledge of the design team, they help in generating and selecting a range of design directions that are well-suited to the target user population and context of use.

## Create and Refine Specific Solutions and Prototypes

### *Objectives*

The goal of phase 3 is to further develop and detail the attributes of selected options to the point that a specific solution can be implemented as a final design. Questions that are commonly addressed in phase 3 will include the following:

- Are the functionality, usability and appeal of the design compatible with what we know of people's needs and desires?
- How rapidly do people understand the potential of the design and ways to interact with it?
- What are the detailed appearance and technical performance requirements that will meet usage needs?

*Example: Your team has established a couple of possible design directions for a wrist-worn medical monitor that people will need to wear 24 h a day. You plan to try out some weighted 3D prototypes/models.*

- How will this feel when you have to wear it 24 h per day?
- How can people cope with it through daily activities like sleeping, dressing, bathing and getting in and out of a car?
- What do we need to change to make it as secure and comfortable as possible?

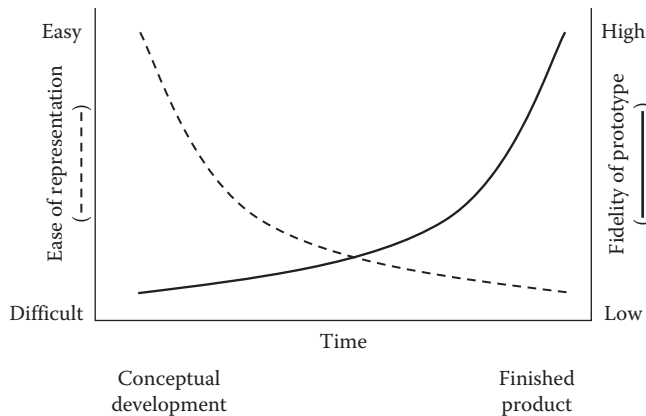
Phase 3 clearly builds on the outcome of phase 2. Phase 2 tends to be a divergent phase in which the design team pushes the boundaries to explore many design possibilities; phase 3 can be characterised as one of convergence and consolidation. It involves detailing the specific attributes that will eventually be incorporated into the design. This will often involve the need to compare different solutions with each other (often formulated through the production of prototypes of varying levels of fidelity) or to check solutions against specific performance requirements that have been set as targets for the design.

### *How People Can Contribute*

In phase 3, people can contribute to the creation of designs as well as evaluate them. Participants can be involved in co-development with the design team in varying stages of the development, especially in exercises aimed at resolving high-level system or product configurations.

A key issue is to provide participants with sufficient information to enable them to provide useful insight. There is almost always a compromise between the ease with which options and contexts can be represented and the level of fidelity of those representations (prototypes). This compromise is schematically shown in Figure 10.1. The challenge for the design researcher is to match the type of questions that can be investigated to the level of design representation available. In the early phases, when design concepts are fairly loose, it is generally easy and inexpensive to represent them in some form of low-fidelity simulation (such as paper prototypes) that people can respond to. As concepts become more robust and detailed, accurate representations become more time-consuming and often more costly to realise. In most cases, participants are quite able to appreciate the nature of a rough prototype and are tolerant of simulations *breaking down*. Furthermore, participants can often be more motivated to help a design team seeing that there is clearly a chance to shape the solution before it is finalised.

The value people provide here is less in deriving details of specific solutions as much as their ability to highlight requirements, prioritise functionality, and reflect users' own priorities in terms of what is of interest in the solution. Their involvement also helps to remind the design team about aspects concerning the reality of usage and context that might easily be overlooked as commercial and technical factors influence the final implementation of the design.



**FIGURE 10.1** Ease and fidelity of design representation.

## Evaluate Implemented Designs in the Real World and against User Requirements

### Objectives

The focus of phase 4 is on investigating implemented designs in use. The purpose is to discover whether the design is being used as was anticipated during phases 2 and 3, and to check on the assumptions made during the course of development.

Questions that are commonly addressed in phase 4 will include the following:

- Is the design being used by the kinds of people and for the purposes expected?
- How valid were the assumptions about when, where and how the design would be used?
- Is there sufficient support for the design in use, in terms of, for example, training, servicing?
- What considerations and applications were overlooked?
- How is the design changing people's behaviour and experiences?

*Example: Your company has recently launched a new mobile phone with an attached data projector.*

- Are people able to use the device as easily as they use their existing smartphone?
- In what circumstances, both personal and work lives, is the projection capability most used?
- What are the most frequently used functions?
- Did we provide the easiest access to these?
- Are there any functions that are hardly ever used?

### How People Can Contribute

Evaluating designs in use usually involves monitoring actual performance and logging usage data, preferably over an extended period of time. People contribute by volunteering information about their behaviour and experience, often in response to proactive requests to them to participate. They also provide information to design researchers via incidental data they generate, such as records of complaints, service and product returns held by the manufacturer or supplier. Information can also be gathered from existing online communities and discussion groups related to specific kinds of products and services.

By revealing specifics about the use of a new design in context, this phase provides valuable input into phase 1 of any subsequent development of the design and related products and services. It also provides the opportunity to examine how the solution impacts on and is impacted by the actual context of use. This is of particular value in the case of new-to-the-world products and services where

inevitably new patterns of behaviour and expectations develop. Consider, for example, how the availability of mobile phones and text messaging services has changed the way people, especially the young, plan and conduct social events from the days that telecommunications were essentially landline based. Such evolving usage patterns create new challenges and opportunities for design, at both detailed and conceptual levels. Awareness of these opportunities and data about their prevalence is a key competitive advantage and offers great possibilities for user advocacy to drive commercial innovation.

## SETTING UP AN INVESTIGATION: GENERAL CONSIDERATIONS

Later in this chapter, we discuss, in some detail, a range of different methods for providing human-centred input to the four main design phases. But there are several basic issues that need to be considered when planning the involvement of people, irrespective of the particular methods that may be employed.

### Ethical Issues

Participants share their time, energy, information and sometimes even intimacies that are of commercial value to developers. It is important to conduct all investigations with the highest degree of respect and consideration for people's time, health, safety and privacy. Participants should be informed about the purpose of their involvement, what is being recorded and what will be done with the data. Sometimes it is necessary that participants are *not* made fully aware of a design researcher's specific focus ahead of time, for example if the knowledge will alter their behaviour, but in these cases, it is generally courteous and tends to build trust to *let them in on the secret* when a session is completed. For those looking for more information, the British Psychological Society provides a very good set of ethical standards and procedures for conducting research with human participants (see <http://www.bps.org.uk/what-we-do/ethics-standards/ethics-standards>). See Chapters 1 and 4 for further discussion of ethics in E/HF research.

### Objectives

Design research projects can usually be designed to provide useful information within the practical constraints of costs and timescales. The basic aim of such an investigation is to improve on estimates based on opinions by obtaining evidence that can guide the design team. Any investigation needs to start with a clear set of objectives. These will depend upon the current phase of the design process, the questions that have arisen and who on the team needs the information. Here are some typical examples of different information needs within a development team:

- *Product planner*: A sporting goods company is considering its strategy for integrating communications technology into products. What sporting situations and kinds of products would benefit from this kind of integration?
- *Interaction designer*: The interface elements and behaviours for a new digital radio system are being developed. What user considerations should we be aware of in migrating a familiar system from analogue to digital format?
- *Industrial designer*: A new remote-control device is being designed for museum visitors to use in engaging with interactive exhibits. What form, factors and materials will be pleasant to handle, easy to keep clean and enable it be stowed, carried and used most easily?
- *Marketing specialist*: A new range of innovative personal entertainment products is being considered. What are the key features that will appeal to the second- and third-generation users?
- *Product manager*: A choice has to be made between two technologies. Will it matter to the user which we choose?

- *Design engineer*: What are the critical dimensions in the layout of a particular workstation?
- *Software engineer*: A particular user group is accustomed to a particular style of interaction. Will a new style be acceptable, and are there any negative transfer effects?
- *Documentation specialist*: How will customers learn or be trained most effectively, and where might particular difficulties with the product occur?
- *Maintenance engineers*: A specific time limit has to be met for the exchange of particular machine components. Will the design meet this requirement?

Many investigations can, and need to, provide answers to more than one question of this kind. In designing an investigation, it is a useful discipline to articulate and prioritise each of the questions to establish appropriate focus.

### Participant Goals, Tasks and Activities

The design researcher needs to work out what activities they need to observe and be clear about how much they need to influence them in order to satisfy the objectives of the investigation. In general, the earlier in a programme of development research is done, the broader the investigation. It also depends on how much the design researcher wishes to diverge and invent, and how much there is a need to converge on a solution that is already fairly well defined. In some cases, such as on-site observations, the design researcher may have no direct control over what participants do but may need to be selective about what aspects they need to observe. In situations where the design researcher is necessarily absent, for example in the case of *diary-keeping studies*, the design researcher has little control and can describe their requirements only in general terms. In others, such as user trials, the design researcher is in the position to instruct participants to carry out specific tasks. The design researcher also needs to consider the relationship between the tasks and activities that are directly served by the design, and those that are complementary in some way. For example, users of automatic teller machines (ATMs) often have to deal with bags and other types of luggage when making cash withdrawals. This can add to anxieties about personal security when using ATMs in airports, railway stations and similar crowded areas. The complementary aspects may not be the subject of the investigation, but they may be very important in ensuring that the local context is adequately taken into account when designing the investigation.

### Location

The appropriate location for a research project depends upon the nature of the information sought and the phase of the design process. Exploring needs and opportunities (phase 1) often involves at least some activity in the context where the design will ultimately be used. Investigating design options, and refining specific design solutions (phase 2), may also be carried out *in context* but is often carried out within artificial or simulated situations, such as a usability laboratory or workshop space, where circumstances can be controlled by the design researcher. But in these controlled situations, it is important to be aware of the extent to which there may be significant contextual factors that are relevant to the design, and to accommodate these as far as possible using props and models to simulate elements where desirable. Given the pragmatic constraints of time and resources when working in industry, design researchers can often make reasonable assumptions based on their own professional experience, experience of real contexts of use and information provided by participants. However, there is a growing conviction amongst practitioners that studies conducted in a real-life context provide opportunities to gain important insights that would not be revealed in the laboratory (see Chamberlain et al., 2012 for an extensive discussion of research *in-the-wild* in relation to interactive systems development).

### Participant Selection

An obvious point, but nevertheless essential, is that people selected to participate should have relevant experience or expertise to be able to provide insightful contributions. The design researcher



needs to profile a range of key stakeholders who will provide useful perspectives on the design. At a high level, there are also several dimensions to consider in deciding whom to involve:

- *The person as user*: a future operator, passenger, visitor who will interact with the design and may represent *typical* or *extreme* use, be novice, experienced or expert
- *The person as an individual human being*: their physical and psychological capacities and limitations, such as age, strength, literacy level, demographic and/or psychographic segment
- *The person as a social being*: their role in relation to the context of use, for example student, teacher, doctor and their domain of expertise and ultimate goal
- *The person as a cultural being*: their world view, lifestyle and values, habits, rituals and expectations based upon national, ethnic, generational, religious or other adopted group membership

These dimensions overlap, and of course within any one person, they interplay at various levels. The participant profile should identify the dimensions that are most relevant to the design and type of investigation that needs to be carried out. In some cases, there may be data to assist in deciding what is relevant (e.g. if a marketing plan has been defined, there is likely to be demographic and/or psychographic data available). But more often than not, the design researcher has to make an informed estimate based on experience and other useful sources of information. Apart from the professional E/HF literature, information sources could include market research, service, sales, application specialists, training departments or occupational health specialists. All may be able to offer useful guidance. Again, assumptions will have to be made which should be agreed with those who requested the investigation.

Once profiles are developed, participants can then be recruited to match them. Often, these profiles will serve as a guide rather than strict criteria for selection. However, if the validity of an investigation relies on participants complying with specific requirements, then they should be appropriately screened before being recruited.

Selected participants do not always need to be *typical* of the eventual user population. Sometimes it is useful to seek out people who characterise the *fringes* in order to help the design team establish the boundary conditions in terms of their needs and requirements (see, e.g., Gilmore, 2002). However, whenever possible or practicable, it is good practice to collect descriptive data from the participants that enable comparison with, and extrapolation to, the user population as a whole. Of course, such comparisons can only be made if the corresponding data describing the user population are available in an appropriate form.

### Participant Numbers

It is difficult to provide general advice on how many participants to involve in a design research investigation. There are methods available for estimating sample sizes in formal experiments, where statistical probabilities are required to examine the role of key variables in an outcome (Collins, 1986). Practitioners in industry often find however that the number of variables relevant to a design-related investigation is so large that such experimental methods are impracticable and unwieldy (and accordingly, much design research favours qualitative methods where such strict control of participants is less of a priority). It is also often the case that statistical information about the variance of human characteristics within a population is not available and best estimates have to be made. In short, this usually results in estimates of participant numbers being very large, sometimes alarmingly so, to achieve even moderate levels of accuracy.

In practice, involvement of very small numbers of, for example, four to six, participants selected to represent a good inclusive range of characteristics and contexts can provide a satisficing and significant amount of information and evidence of issues that were not apparent beforehand. Many design researchers aim to involve small numbers of participants representing as broad a range of

perspectives as possible on an incremental basis (sometimes referred to as maximum variation sampling). Thus, the design researcher stops recruiting new participants or running new sessions when they are confident that no significant new knowledge is likely to be generated.

For user trials specifically, experience shows that dependable results can be obtained from as few as five users in a single trial (Rubinstein and Hersh, 1984). More than five users will generally always be beneficial, but how many more is an open question. Virzi (1992) concludes that 80% of usability problems can be detected by four to five participants, and fewer new insights are revealed as the numbers increase, with the first one or two participants detecting the most severe problems. These conclusions reflect a widely held view amongst practitioners that small numbers are often sufficient to guide the design team. However, there are many examples where investigations have involved more than 30 users, and in some cases, samples in excess of 100 are quoted. However, always bear in mind that, for example, a visit to even a single surgical procedure and subsequent discussion with the one attending physician and anaesthetist will not necessarily provide a repeatable or representative *result*. But it beats relying solely on a discussion around the interaction designer's computer about how the blood–gas analyser interface should be laid out.

Recent advances in web connectivity have meant that there are new ways of including vast numbers of potential users in the design process (albeit largely in the earlier phases we have discussed). Techniques such as crowdsourcing (Brabham, 2008) take large tasks and break them down into smaller activities, which are then made available to the public at large to complete, for micro-payment. Some design companies are beginning to explore how such techniques might be fruitfully utilised specifically for design research to engage with significant numbers of research participants around the world (see, e.g., IDEO's OpenIDEO <http://www.openideo.com> and Frog Design's frogMob <http://frogmob.frogdesign.com>).

## Representing the Design

It is often necessary to represent the design in some way in order to undertake an investigation. The appropriate level of fidelity (from low to high) will influence the type of information required and vice versa (see Figure 10.1). Many types of representation are commonly used, including the following:

- *'Paper-based' descriptions of product concepts*: These are interaction sketches, storyboards, annotated drawings, wireframes, screen graphics or other concept descriptions that can enable initial explorations of ideas on product functionality to be made, important usability characteristics to be identified or *walkthrough* studies of protocols for product interfaces to be conducted (see Buxton et al., 2011).
- *Part prototypes or simulations*: Part prototypes are used to simulate specific functional attributes of a design. They might be mock-ups of physical form, scale or mass, mechanical models, static or animated screen graphics that enable people to interact with them. The prototype may look nothing like the final design but will accurately represent those aspects under investigation. Part prototyping of software through the use of rapid prototyping tools is one area where this type of testing is now very common.
- *'Experience' prototypes*: These are representations in any medium that help people to appreciate experiential issues beyond the purely functional attributes of a design. They are designed to include contextual and affective qualities conveyed through a relevant subjective experience (see Buchenau and Fulton Suri, 2000).
- *Full prototypes*: Full prototypes perform as the final product is intended to perform and incorporate the complete functionality and appearance of the product.
- *Complete products*: Complete products enable the complete user interface to be examined. This opens up the possibility of carrying out field investigations, comparative studies with other products, in-service studies, etc. For further readings about design representations particularly in HCI, see Houde and Hill (1997) and Wong (1992).

## Measures, Observations and Subjective Data

Some investigations will reveal simple objective data, such as error rates or times to complete a transaction. Others will enable the design researcher to make direct objective observations of behaviour and events. Still others will provide subjective information in the form of participant's expressions and self-reports of behaviour, opinions, attitudes, concerns and habits (see Chapter 1 for discussion of objective/subjective methods and Chapter 4 for further discussion of self-report methods).

*Objective measurement* has traditionally been a central tenet of human factors research. Kantowitz (1992) summarises its role: 'Measures are the gears and cogs that make empirical research efforts run. No empirical study can be better than its selected measures'. In the design context, the standard empirical approach is time-consuming and appropriate for only some of the issues that need to be addressed. Nevertheless, it is wise to obtain objective data whenever it is possible and practicable. Such data are readily accepted by many individuals (especially those with a more traditional *engineering* mindset typical of many technology organisations). It can complement subjective and qualitative data, and insights gathered from other more exploratory methods with participants. There are three main kinds of objective data that are frequently collected:

- Direct measurements taken from a person. Examples include body dimensions and physiological measurements such as heart rate, oxygen intake and body temperature.
- Data resulting from user actions recorded by the design researcher or by some remote means such as video or automatic event logging such as keystroke or touch-screen event capture. Examples include time-based measures (e.g. task duration, event duration, response time, reaction time) and error or accuracy scores (e.g. mistakes in procedures, incorrect responses to stimuli, error rates in relation to time or events).
- Data taken directly from product or prototype as the result of adjustments or modifications by participants. Examples include positions for seats, shelves or controls, levels for lighting, sound, colour, quality, brightness and contrast settings on displays.

*Observations of user actions* (see also Chapters 4, 6 and 13) also fall into three main categories (adapted from Meister, 1986):

- Descriptive techniques, where the observer simply records events as they take place (e.g. time based, frequency based, event sequence, postures adopted and controls used)
- Evaluative techniques, where the observer evaluates the outcome or consequence of events that have taken place (e.g. degrees of difficulty, incidence of hazardous events and errors of judgement)
- Diagnostic techniques, where the observer identifies the causes that give rise to the observed events (e.g. positioning of controls, inadequacy of displays and poor user instructions)

Descriptive observations are generally the easiest to set up and conduct. The degree of difficulty increases with evaluative observations and is greatest with diagnostic observations. This is primarily because of the type of on-the-spot judgements the design researcher is required to make during the observation. Ability to conduct evaluative or diagnostic observations depends greatly on the background knowledge of the design researcher. To make effective evaluative or diagnostic observations, the design researcher needs extensive background knowledge of tasks and objectives. It is essential also to decide what to look at and what to record beforehand. The greatest problems come from the speed and number of events occurring concurrently. It often surprises people who are

new to observing the behaviour of people just how quickly the accuracy and reliability of a design researcher can break down when there are too many events to monitor. Even moderately experienced design researchers often find themselves being far too ambitious in the variety of events they wish to record. Consequently, it is most important that the method of observation is carefully piloted and video recording used as back-up whenever possible.

*Subjective data* such as participant's self-reported behaviour, opinions, and habits are often difficult to quantify in any meaningful way (see Chapter 4). However, they are frequently the basis for rich insights that can be applied to design and should be carefully documented and represented. These data often find their way into *stories* told about specific people observed, their behaviour and expressions. Such stories have an important role in carrying information from investigations to design teams. Design researchers bringing *tales from the field* – narrative descriptions and anecdotes highlighting specific examples derived from both objective data and observations – help to bring data to life, make them actionable and make connections for other team members who may find charts and tables of data irrelevant. Stories are inherently not objective, but are selectively edited and integrated descriptions of phenomena. Care must be taken in the choice and telling of such tales that they maintain the integrity of their source and reflect significant design issues in appropriate ways.

### Gathering Behavioural Data

There are many techniques for gathering behavioural data. The one technique we want to highlight here is the video. In recent years, video technology has become the most prolific method for recording participant behaviour. Video not only has become a standard equipment in any usability laboratory (Nielsen, 1994) but also, due to the portability of modern cameras and the excellence of the recorded images, is recorded much more frequently in less formal on-site studies. The great advantage of video is that it readily captures concurrent events, and of course, it enables one to repeatedly revisit the data. A particular sequence of events can be analysed several times over from different viewpoints, enabling a design researcher to overcome many of the problems associated with analysing events which occur in parallel or very quickly. A well-known example of this was a study of London Underground Control Rooms by Heath and Luff (1991), in which fine-grained video analysis revealed the complex and critical practices of coordination and peripheral situational awareness of activity between users of different workstations within the control room. However, a few cautionary words are in order. First, detailed analysis of video recordings can be extremely time-consuming. Ratios of at least between 1:5 and 1:10 (*recorded time: analysis time*) are commonly quoted (Mackay et al., 1988); Bainbridge and Sanderson (1995) report *analysis time: sequence time* ratios of 10:1 to 1000:1 (although there are an increasing number of technology-aided approaches to support analysis becoming available which have the potential to significantly reduce these ratios). Second, the need for detailed preparatory work is not reduced because an extensive record of events is available. The use of video requires at least as much care as paper-based approaches in deciding beforehand what events are to be observed and the form in which the data are to be analysed. Detailed discussions of video analysis techniques (often referred to as *Interaction Analysis*) can be seen in Jordan and Henderson (1995) and Heath et al. (2010).

### Interpretation of Results

No matter what kind of information emerges from investigations involving users, it requires careful interpretation in order to pull out applicable insights. Rarely will an investigation in itself provide information that can be literally applied to a design. Rather, the investigation will first provide broad insight into people's current perceptions and sensitivities and help to establish a frame of reference from which the design team can think about future offerings. Second, the investigation

will illuminate issues that can be turned into specific design criteria and design principles that need to be met if the design is to be successful. These criteria will guide design efforts and serve as a reference for the evaluation of proposals. There are many different analytic orientations that a design researcher might adopt when analysing data; Randall et al. (2008) provide a good discussion of several important approaches and their relevant merits, with regard to the development of interactive systems.

### Design Criteria

Deciding on which criteria to use to guide the design process and evaluate a design is key. In some cases, the appropriate criteria will be self-evident, but defining them is often the specific focus of investigations with users, particularly when the technologies are new and usage patterns are still evolving; see, for example, Dillon (1992).

The ultimate success of a design may be related to diverse issues including social, commercial, organisational, subjective, emotional, value for money concerns in addition to those physical, cognitive, safety, reliability and usability criteria that are more usually considered. The goal of defining design criteria is to define those dimensions that will correlate with a successful design, be measurable, and produce information that facilitates design decisions.

Criteria may be phrased as, for example, ‘the haircutting tool should be quiet enough to allow stylist and client to hold a conversation’ or ‘the voting transaction should feel reassuring’. Even for these, it would ultimately be possible to develop some objective criteria (e.g. decibel and speech interference levels, system response and speed). In practice though, these criteria may be more effectively judged subjectively by participants and design team members as a design evolves (Fulton Suri, 2001). In any case, each criterion needs to be expressed in such a way that the design, or a participant’s performance with it, can be assessed, whether subjectively or objectively. It is usual practice to use several measures, rather than to rely on one, and to include both objective and subjective measures.

### METHODS: OVERVIEW OF AREAS

We have clustered the methods for involving people in design into a number of categories. Within each category, we discuss the general approach and refer to specific examples used by different design researchers. We do not pretend to provide a comprehensive inventory of all examples of each method. That would require a book in itself. As said in the introduction, the purpose of this chapter is to raise awareness of the scope of methods available.

Many of the methods outlined are about how the design researcher can create the right kind of circumstances that help participants contribute and help the design team to create the appropriate design. On the one hand, good methods should help participants to articulate their needs, wishes, experiences and expectations, and provoke their imagination. On the other hand, good methods should enable the design researcher (and the design team) to generate the most useful information that helps the design team make the best decisions.

There are many books of methods for HCD, with varying levels of depth. Good overviews of techniques can be found in Hanington and Martin (2012), Kuniavsky et al. (2012) and Lazar et al. (2009). Overviews of many methods can also be found at the Usability Net website (<http://www.usabilitynet.org/tools/methods.htm>).

We have rated each method according to its ability to deliver the type of information required at each design phase. The rating is meant as only a guide and is based on a combination of our professional experience and comments by design researchers on the methods in case study examples.

| Methods                        | References                                                            | Explore and Understand Needs | Identify Design Options and Directions | Create and Refine Specific Solutions | Evaluate Designs in Real-World Situations |
|--------------------------------|-----------------------------------------------------------------------|------------------------------|----------------------------------------|--------------------------------------|-------------------------------------------|
| <i>Contextual observations</i> |                                                                       |                              |                                        |                                      |                                           |
| Behavioural archaeology        | IDEO (2003)                                                           | ☺☺☺                          | ☺☺                                     | NA                                   | ☺☺                                        |
| Behaviour mapping              | IDEO (2003)                                                           | ☺☺☺                          | ☺☺                                     | ☺                                    | ☺☺☺                                       |
| Blueprint mapping              | Muller et al. (1997)                                                  | ☺☺☺                          | ☺☺                                     | ☺                                    | NA                                        |
| Cultural probes                | Gaver et al. (2004)                                                   | ☺☺☺                          | ☺☺                                     | NA                                   | NA                                        |
| Day in a life                  | IDEO (2003)                                                           | ☺☺☺                          | ☺☺                                     | NA                                   | ☺☺                                        |
| Contextual enquiry             | Holtzblatt et al. (2004)                                              | ☺☺☺                          | ☺☺                                     | NA                                   | ☺☺                                        |
| Ethnography; rapid/video       | Crabtree et al. (2012), Randall et al. (2008)                         | ☺☺☺                          | ☺                                      | NA                                   | ☺☺☺                                       |
| Guided tours                   | Kirk and Sellen (2010), Pettrelli et al. (2008)                       | ☺☺☺                          | ☺☺                                     | ☺                                    | ☺☺☺                                       |
| Narration                      | IDEO (2003)                                                           | ☺☺☺                          | ☺☺☺                                    | ☺☺☺                                  | ☺☺☺                                       |
| Shadowing                      | IDEO (2003)                                                           | ☺☺☺                          | ☺☺☺                                    | NA                                   | ☺☺☺                                       |
| <i>Diary-keeping</i>           |                                                                       |                              |                                        |                                      |                                           |
| Audio/video/written diaries    | Lindley et al. (2012), Palen and Salzman (2002), Sellen et al. (2002) | ☺☺☺                          | ☺                                      | NA                                   | ☺☺☺                                       |
| Narration                      | IDEO (2003)                                                           | ☺☺☺                          | ☺☺☺                                    | ☺☺☺                                  | ☺☺☺                                       |
| User photo-surveys             | IDEO (2003)                                                           | ☺☺☺                          | ☺                                      | NA                                   | ☺☺☺                                       |
| <i>Framework development</i>   |                                                                       |                              |                                        |                                      |                                           |
| Affinity diagramming           | Harboe et al. (2012)                                                  | ☺                            | ☺☺☺                                    | ☺☺☺                                  | NA                                        |
| Brain draw                     | Muller et al. (1997)                                                  | ☺☺☺                          | ☺☺☺                                    | NA                                   | NA                                        |
| Cluster and network diagrams   | Tufte (1990)                                                          | ☺☺☺                          | ☺☺☺                                    | ☺                                    | NA                                        |
| User modelling                 | Hasdogan (1996)                                                       | ☺☺☺                          | ☺☺☺                                    | ☺☺                                   | ☺☺                                        |
| Journey mapping                |                                                                       | ☺☺☺                          | ☺☺                                     | NA                                   | NA                                        |
| Task-flow schematics           |                                                                       | ☺                            | ☺☺☺                                    | ☺☺☺                                  | ☺☺                                        |
| <i>Interviews</i>              |                                                                       |                              |                                        |                                      |                                           |
| Chat rooms                     |                                                                       | ☺☺                           | ☺☺                                     | NA                                   | ☺                                         |
| Focus groups                   | Langford and McDonagh (2003)                                          | ☺☺☺                          | ☺☺☺                                    | ☺                                    | ☺☺☺                                       |
| Interviews (one-on-one)        | Oppenheim (1999)                                                      | ☺☺☺                          | ☺☺☺                                    | ☺☺☺                                  | ☺☺☺                                       |
| Interviews (small group)       | Langford and McDonagh (2003)                                          | ☺☺☺                          | ☺☺☺                                    | ☺☺☺                                  | ☺☺☺                                       |
| <i>Questionnaires</i>          |                                                                       |                              |                                        |                                      |                                           |
| Questionnaires (in person)     | Oppenheim (1999)                                                      | ☺☺☺                          | ☺☺☺                                    | ☺☺☺                                  | ☺☺☺                                       |
| Questionnaires (online)        | Poynter (2010)                                                        | ☺☺☺                          | ☺☺                                     | ☺                                    | ☺☺☺                                       |
| Questionnaires (survey)        |                                                                       | ☺☺☺                          | ☺☺                                     | ☺                                    | ☺☺☺                                       |
| <i>Projective techniques</i>   |                                                                       |                              |                                        |                                      |                                           |
| Brain draw                     | Muller et al. (1997)                                                  | ☺☺☺                          | ☺☺☺                                    | ☺                                    | NA                                        |
| Collage-making                 | Sanders and William (2001)                                            | ☺☺☺                          | ☺☺                                     | NA                                   | NA                                        |
| Lego Serious Play              | Cantoni et al. (2009)                                                 | ☺☺☺                          | ☺☺☺                                    | NA                                   | NA                                        |
| Sketching                      | Buxton et al. 2011                                                    | ☺☺☺                          | ☺☺                                     | NA                                   | NA                                        |

(Continued)



| Methods                                      | References                                  | Explore and Understand Needs | Identify Design Options and Directions | Create and Refine Specific Solutions | Evaluate Designs in Real-World Situations |
|----------------------------------------------|---------------------------------------------|------------------------------|----------------------------------------|--------------------------------------|-------------------------------------------|
| <i>Role-playing</i>                          |                                             |                              |                                        |                                      |                                           |
| Artefact walkthrough                         | Muller et al. (1997)                        | 😊😊😊                          | 😊😊😊                                    | 😊😊😊                                  | 😊😊😊                                       |
| Body-storming                                | Buchenau and Fulton Suri (2000)             | 😊😊😊                          | 😊😊😊                                    | 😊😊                                   | NA                                        |
| Cognitive walkthrough                        | Spencer (2000)                              | NA                           | 😊😊                                     | 😊😊😊                                  | 😊😊😊                                       |
| Place storming                               | Anderson and McGonigal (2004)               | 😊😊😊                          | 😊😊😊                                    | 😊😊                                   | NA                                        |
| Role-playing                                 | Boess et al. (2007)                         | 😊😊😊                          | 😊😊😊                                    | 😊😊                                   | NA                                        |
| <i>Scenario building</i>                     |                                             |                              |                                        |                                      |                                           |
| Design fiction                               | Blythe and Wright (2006), Sterling (2009)   | 😊😊😊                          | 😊😊                                     | NA                                   | NA                                        |
| Personas                                     | Adlin and Pruitt (2010)                     | 😊                            | 😊😊😊                                    | 😊😊😊                                  | 😊                                         |
| Storyboards                                  | Verplank et al. (1993)                      | 😊                            | 😊😊😊                                    | 😊😊😊                                  | NA                                        |
| Story telling                                | Joe (1997), Moggridge (1993)                | 😊😊😊                          | 😊😊😊                                    | 😊                                    | 😊😊                                        |
| Task-flow schematics                         | Holtzblatt et al. (2004)                    | 😊                            | 😊😊                                     | 😊😊😊                                  | 😊😊😊                                       |
| Usage scenarios                              | Carroll (1995), Rosson and Carroll (2002)   | 😊😊😊                          | 😊😊😊                                    | 😊😊😊                                  | 😊                                         |
| Video prototyping                            | Briggs et al. (2012), Mancini et al. (2010) | 😊😊😊                          | 😊😊                                     | NA                                   | NA                                        |
| <i>User trials</i>                           |                                             |                              |                                        |                                      |                                           |
| Acceptance tests                             | Cantwell and Stajano (1985)                 | 😊                            | 😊😊                                     | 😊😊😊                                  | 😊😊😊                                       |
| Co-discovery                                 | Kemp and Gelderen (1996)                    | NA                           | 😊😊                                     | 😊😊😊                                  | 😊😊                                        |
| Cooperative evaluation                       | Monk et al. (1993)                          | 😊😊😊                          | 😊😊😊                                    | 😊😊😊                                  | 😊😊😊                                       |
| User performance trials                      | Fulton Suri (1993)                          | 😊                            | 😊😊                                     | 😊😊😊                                  | 😊😊😊                                       |
| Usability tests                              | Faulkner (2000)                             | NA                           | 😊😊                                     | 😊😊😊                                  | 😊😊                                        |
| Verbal and think-aloud protocols             | Olmsted-Hawala et al. (2010)                | 😊😊                           | 😊😊😊                                    | 😊😊😊                                  | 😊😊😊                                       |
| Wizard of Oz                                 | Bradley et al. (2010)                       | 😊                            | 😊                                      | 😊😊😊                                  | NA                                        |
| <i>User workshops</i>                        |                                             |                              |                                        |                                      |                                           |
| Activity groups                              | Druin (2002)                                | 😊😊😊                          | 😊😊                                     | NA                                   | 😊                                         |
| Collaborative/participatory design workshops | Pedersen and Buur (2000)                    | 😊                            | 😊😊😊                                    | 😊😊😊                                  | NA                                        |
| Cooperative requirements capture             | Muller et al. (1995)                        | 😊😊😊                          | 😊😊😊                                    | 😊                                    | 😊                                         |
| Future workshop                              | Kensing and Madsen (1991)                   | 😊😊😊                          | 😊😊                                     | 😊                                    | 😊                                         |
| Storyboards                                  | Muller et al. (1995)                        | 😊                            | 😊😊😊                                    | 😊😊😊                                  | NA                                        |
| Story telling                                | Moggridge (1993)                            | 😊😊😊                          | 😊😊                                     | 😊                                    | NA                                        |

😊😊😊, strong; 😊😊, neutral; 😊, weak; NA, not applicable.

## CONTEXTUAL OBSERVATIONS

### OVERVIEW

Contextual observations, sometimes termed ethnographic methods, include a range of methods that involve learning about people's behaviour and activities as they occur in a real-world setting, rather than in a controlled environment. They are an excellent way to see how technologies, products and



artefacts are used in practice, and how the real conditions of everyday interaction impact on the way they are used. Applications vary in breadth of focus, in the degree to which the design researcher influences and interacts while collecting information, and in an extent to which current rather than historical events are of interest. The focus may be upon a very limited set of behaviours, for example how people empty vacuum cleaners, or upon a broader domain, for example how are house-cleaning activities performed. Observations may involve minimal disruption to natural and real-time behaviour, for example watching without interrupting or video-recording using a discreetly positioned time-lapse camera. Alternatively, they may rely heavily on a participant's involvement with the design researcher, as in shadowing an individual over a period of time. Others depend to a higher degree upon recall of past or typical ways of behaving. For example, participants may be asked to demonstrate, explain and/or reflect upon how they perform a particular task, or to act as a tour-guide and interpreter for the design researcher, pointing out and commenting on the significance of artefacts and design elements in an environment. The design researcher may be studying the use of technologies in general as they exist, evaluating specific existing designs or advanced design prototypes in context. Two variations on this theme are behaviour sampling and shadowing.

### **BEHAVIOUR SAMPLING**

Behaviour sampling involves making a series of brief, sometimes momentary, observations of people's activities, often over an extended period of time and in a variety of locations. It is helpful in situations where variation over time of day, week and month are of interest, but it is not practical to maintain continuous observation. Behaviour sampling can be achieved by remote monitoring through time-lapse video-recording or by a design researcher making personal visits, for example, to identify the locations within a hospital that are congested or underused and where staff and patients tend to congregate or seek private moments. Behaviour sampling is also sometimes achieved by enlisting participants' help remotely. An example is described in Fulton Suri (2001) in which people were given pagers and called intermittently over a 1-week period and asked to record where they were, what they were doing and with whom. The goal was to anticipate the needs and fears of future patients with implanted defibrillators that might shock them at any time as they go about their daily lives.

### **SHADOWING**

Shadowing is a form of contextual observation that involves following selected individuals as they conduct everyday activities through space and over time. This is a useful method for examining, for example, how specific work-roles play out in practice, as individuals interact with other people, tools, technologies and processes in the conduct of their working day. It is also a helpful way to understand the relative delights and frustrations that people experience as they interact with products and systems over time. In the context of designing hospital systems, products and services, for example, it may be appropriate to shadow a sample of caregivers, including doctors and nurses, as well as a sample of patients and their loved ones to elucidate the range of different behaviours, perceptions and needs.

### **WHEN TO USE**

Contextual observations can be used at all stages of the design process, but they are often most valuable in the early stages when a design team needs to develop an understanding of user needs, requirements and preferences. This may be particularly important when the design team is not familiar with the usage context. Contextual observations can reveal

- Physical and social environmental factors that need to be taken into account
- People's actual practices in contrast with presumed or normative descriptions

- Insights concerning user preferences and attitudes, and the impact of social–cultural issues
- Workarounds developed to cope with design inadequacies and/or people’s own physical/cognitive abilities
- Functional and emotional significances of specific artefacts or rituals
- Unusual physical or cognitive demands where dynamic behaviour in space and temporal factors may need to be accommodated
- Opportunities for design improvements in terms of people’s processes and experience

Contextual observations can provide realistic (and often critical) contextual texture for the design team to use as reference material throughout the design process.

## **PARTICIPANT INVOLVEMENT**

Participants may be passively or actively involved in collecting information through these methods. Where participants are only passively involved, it is helpful to have at least one or two participant representatives to assist with interpreting observed data or to triangulate using other methods such as debrief interviews. Active involvement of a design researcher with participants can create an empathic relationship, which can lead to unexpected insights, but also to undue influence over participant behaviour. The appropriate relationship will depend in part on whether you are really seeking insight and inspiration, or factual data. It is generally best not to maintain a formal tone. Aim for a conversation rather than an interview to maintain naturalism. Long periods of immersion are often better, especially in situations where time is needed to settle down into the *culture*.

## **DESIGNER AND DEVELOPER INVOLVEMENT**

It is recommended that, as far as possible, the design team is directly involved in fieldwork and subsequent analysis. Designers and developers can make observations themselves, or at least review raw recorded materials, so that all design team members get first-hand exposure to the realities of how the context might affect the design.

## **HOW TO USE**

### **Process Involved**

Decide on the contexts and vantage points (static or dynamic) that are needed. Decide how participants will be involved and to what extent the design researcher needs to influence or participate with them. The design researcher needs to take care that they get to see what is relevant to their enquiry and that enough time is taken to ensure that the required information is collected. Develop a checklist of things the design team needs to know about and decide on the kind of data collection to be used. The design researcher also needs to take care to get any *permission* to observe and/or get access to the context of interest. Prepare the way in which the data will be analysed and reported. As with all investigations, pilot your data collection techniques and your analysis methods first.

### **Expertise Needed**

If active collaboration with participants is required, the design researcher needs to quickly embed themselves in the local culture and establish an effective rapport with the participants. Parameters to consider may include language, age and/or experience depending on the nature of the enquiry.

## **OUTCOMES**

The reporting of discoveries derived from contextual observations is most valuable when it preserves the rich audio/visual data that are in itself informative and inspiring to the design team.

Raw data are frequently in the form of audio/visual records, whether still-images, video, sound recordings, maps, plans and sketches. The goal of contextual observations is often to get a holistic view of the usage circumstance of interest. Consequently, contextual observations can often deliver large quantities of very rich and diverse data, which can be difficult to analyse. The process of design requires the observations to be interpreted in terms of how the design needs to accommodate the *lessons learnt*. This challenges the design researcher to draw out useful information for the design team quickly and efficiently. The design researcher will need to apply structure to get at high-level issues and more generally applicable principles. So it is often best to report back in debrief presentations or workshops that incorporate video-clips and photographic imagery to illustrate the general principles.

## EXAMPLES OF USE

Eggen et al. (2003) report on a study into the home in which contextual observations featured as a method for eliciting peoples' perceptions about what made a *house* a *home*. Gaver and Dunne (1999) report on a study into the perceptions of elderly people to the use of technologies in a large housing estate. Kirk and Sellen (2010) report on the use of guided tours in homes to explore the design of memory storage and archiving technologies for domestic spaces. Tolmie et al. (2002) report on the use of ethnographic methods to investigate domestic routines as part of a European Union project on the application of ubiquitous computing. Fulton and Stroud (1981) describe the use of fixed cameras and time-lapse recording of passengers on the London Underground and in other locations to capture the incidence of key behaviours. Mäkelä and Fulton Suri (2001) describe a design case study involving the invention of computer-based toys for young children.

## DIARY-KEEPING

### OVERVIEW

Diary-keeping methods include a range of self-documentary techniques in which participants are asked to record events, thoughts and feelings through the course of specific activities or time periods within their natural context. The key feature of diary-keeping is that a participant is not directly supervised by a design researcher. The methods involve asking participants to make written, web-based, photographic, video or audio recordings that can be later reviewed by the design researcher, with or without further involvement of the participant. Depending on exactly what you ask people to do and record, you can gather information about contexts of use, activity through time and space. Diaries can also be used to collect data on preferences, frustrations, needs, motivations and attitudes (see Chapter 4 for further consideration of diary approaches).

### WHEN TO USE

Diary-keeping is valuable at all stages of the design process, though most useful during initial explorations and where prototype solutions are available for evaluation in context. Diary-keeping also allows examination of how people are really using and reacting to the design once it is implemented. Diary-keeping methods allow you to gather information from the context of use without having to be there physically. This is useful when you need to know about behaviour taking place over time and in places that are difficult to observe directly. It also means that you can gather information from geographically/culturally diverse contexts fairly inexpensively. Diary-keeping can be useful as a starting point for follow-up interviews or contextual observations. For example, to learn more about specific events or interactions of interest. In particular, methods that involve visual records can serve as a useful focus for interviews with people where there might be initial reluctance

or difficulty in just talking as, for example, with children or people from another culture. Diary-keeping can also be used as a form of behaviour-sampling technique. For example, participants can be prompted via phone call, text, email or IM at random times of day to capture thoughts, activities and context instantaneously. This allows you to gather more objective data as it is rather less dependent upon self-editing. In some cases, it works well to ask participants to be involved in diary-keeping as a follow-up to interviews or other activities. They may be more motivated to participate after they have made a connection to the subject matter and the design researcher.

### **PARTICIPANT INVOLVEMENT**

Participants are obviously actively involved in collecting the information required. These methods put particular emphasis on ensuring that instructions are clear and that the data can be easily collected. It may be advisable to give participants some training in the type of data the enquiry requires as part of the briefing session.

### **DESIGNER AND DEVELOPER INVOLVEMENT**

By definition, designers and developers have no involvement in the collection of data from participants. However, it can be useful to involve designers and developers in the diary-keeping exercise as respondents themselves; their participation can heighten their general awareness and sensitivity to user issues, and whet their appetite for information obtained from users with experiences similar and dissimilar to their own. As with other methods, close involvement with the preparation of the questions, the design of the data gathering tools and reviewing the results is advised, with potential interim review during the investigation to allow scope for influencing the responding style/content of the participants.

### **HOW TO USE**

#### **Process Involved**

To get the most value out of these methods, you need to provide participants with equipment and instructions that make it very easy to do what you want them to. It is best to design and create self-contained *kits* that include everything they will need. A kit might include instructions about what and when you want them to record, a logbook with defined fields to check or fill in, prompts to remind them to think of specific issues, labels/stickers to attach to photographs they take, camera/video/tape recorder, instructions about how to use the equipment, contact information and instructions about returning their contributions.

#### **Expertise Needed**

The key to designing and implementing good diary-keeping methods is to make it simple and enjoyable for participants to provide the information you need. Like questionnaires, diary-keeping tools usually benefit from careful design of graphic layout so that experienced graphic designers can be a great help. Data are generally qualitative and so benefit from being analysed sensitively, by a team, experienced in this kind of data analysis.

### **OUTCOMES**

As with all self-reporting methods, what you learn depends upon people's motivation, ability and willingness to record information. It can be difficult to control the quality of the data recorded and to ensure that the design researcher receives a complete record. Design researchers can often be faced with big differences in the level of detail and the accuracy of the data recorded. In some cases, this might be the information the enquiry requires, but if not then, the only defence is to carefully

prepare the participants. It is often valuable to explore issues using complementary methods such as debrief interviews, focus groups and user workshops.

## EXAMPLES OF USE

Sun et al. (2013) present three case studies of mobile devices as convenient tools for supporting diary studies, critically examining the design of mobile phone-based diary methods.

Lindley et al. (2012) report on a study of diaries kept of web searching to help develop next-generation search tools.

Palen and Salzman (2002) review many different diary-keeping techniques in the context of studying mobile work.

Sellen et al. (2002) report on a diary study of how knowledge workers use the web.

## FRAMEWORK DEVELOPMENT

### OVERVIEW

Frameworks are interpretive tools for a design team. They help build consensus and make sense of complex information around user issues. They comprise simple graphical models and diagrams that enable the team to structure information and provide a shareable view of key issues and inter-relationships. Their purpose is to make these issues and relationships explicit in a way that helps the design team make their design decisions. Framework development involves many different ways of representing ideas, discoveries or knowledge about people's behaviours, activities, thoughts and feelings in simplified summary form. The focus may be upon the relationship of behaviours through time or space, for example a sequence of events or actions involved in a specific task or experience – these are sometimes referred to as task synthesis or journey maps. Alternatively, the focus may be on more abstract similarities and differences between types of activities, motivations, people or contexts. For example, in designing a new airline check-in system, one simple relevant framework might be a diagram sequencing information flow between a user and the system; another might be a matrix differentiating users as experienced vs. inexperienced and resident vs. foreign visitors. In many cases, framework development is a matter of making explicit and agreeing upon what are actually implicit conceptual models amongst members of the design team or user community. Frameworks may be formalised descriptions as in traditional flow charts describing task sequences, or looser more informal visual descriptions that suggest relationships, such as affinity diagrams or network maps.

### WHEN TO USE

Framework development is a useful synthesis activity that is applicable to all kinds of investigative methods. The activity of creating frameworks enables the design team, with or without the involvement of participants, to discuss and agree upon useful patterns that capture otherwise complex details. The greatest value is in creating an explicit and shared conceptual model for the team to reference. So, in other words, frameworks can be powerful stimulants of communication primarily within the design team, but also between design team and participants. Framework development is valuable for different reasons in all phases of the design process identified earlier.

- In phase 1, it is useful to develop *frameworks for looking* to focus attention and help decide on specific activities/issues/people to involve in explorations. It is useful, following explorations, to develop *frameworks for explaining* that summarise the discoveries made and areas where more information might be necessary.

- In phase 2, the same frameworks can be used to structure opportunity areas and provide a basis for the generation of design ideas – *frameworks for idea generation*. These frameworks also serve as a simple tool to communicate why a team has focussed on a specific set of design aspects.
- In phase 3, frameworks serve as a reference for planning evaluation studies and highlighting where problems arise.
- Phase 4 to focus the enquiry on the key issues that need to be evaluated.

## **PARTICIPANT INVOLVEMENT**

Users may be passively or actively involved in the development of frameworks. Users are passively involved when they implicitly provide the data that underpin the framework. In such cases, it is important to check the validity of the framework by reference to specific individuals who have been, or will be, involved in first-hand enquiries, for example contextual observations and/or interviews. When participants are actively involved in the development of a framework, they might be asked to contribute their own conceptual maps and then together with the design team and a facilitator to develop a consensus view (i.e. being active participants in the construction of themes within an affinity diagram).

## **DESIGNER AND DEVELOPER INVOLVEMENT**

As far as possible, the design team should be involved directly in developing frameworks. All team members should be encouraged to surface their own preconceptions and together reach understanding and take ownership of the framework.

## **HOW TO USE**

### **Process Involved**

As described earlier, frameworks can be used for different purposes. Although sometimes the same basic framework can be used at any phase, it is often useful to consider alternative models. Developing useful frameworks is a creative activity. The way frameworks evolve is highly influenced by the nature of a design project, the phase the design team is in and the type of material the design team wants to describe. In essence, it is about finding and representing graphically the relationships that exist between elements. Elements might be ideas, individual people, types of people, activities, tasks, goals, physical things, spaces, etc. One good way to start is simply to list these elements and look for clusters, sequences, connections and patterns that are interesting. There is no standard way to create a useful/relevant framework although there are a couple of fail-safe ways to proceed. One is to use the intrinsic spatial and temporal dimensions associated with interactions, such as the steps involved in a process (e.g. sending a text message or being the victim of a medical emergency), or the links between people and spatial elements (e.g. types of communications between people in a work setting). Another is to look for polar opposites that describe important dimensions, for example personal – shared, specific – generic, technophobic – technophilic, high investment – low investment, etc.

### **Expertise Needed**

Framework development benefits from the involvement of a combination of analytical and visual thinkers who are also able to help others interpret their own ideas or points of view in a visual way. People skilled in visualisation of information can also help maintain the pace of the process by quickly articulating structure and relationships between elements.



## OUTCOMES

Frameworks and diagrams are helpful in providing an easy way of understanding complex information. They can be very useful not only in clarifying issues for a design team but also for presenting the demands and challenges of a project to clients and senior management. They very often reveal the relative strengths of relationships, gaps in knowledge, opportunities for improvement or innovation and so on. Frameworks such as task-flow diagrams, for example, can reveal where delays or excessive mental workload and hence errors might occur. Network diagrams will show up key formal/informal communication channels and social structures that the design must support. Matrices of people types, activities and/or contexts can show up areas or parts of the process that have been neglected.

## INTERVIEWS

### OVERVIEW

Interviews are an inherent part of involving participants in the design process. Interviews are an essential and valuable method of data collection for all phases of the design process. Interviews are basically a guided conversation between a respondent and a design researcher about his or her perspective on some issue of mutual interest from which the design researcher wants to collate certain information. Interviews may be highly structured and focussed on specific questions, they may be open free-ranging discussions or they may be any one of several combinations. Unfortunately, there has been little discussion in the literature on the strengths and weaknesses of interview techniques in design development work and more particularly any systematic evaluation of the approaches (McClelland, 1984; Meister, 1986). However, most experience with interview techniques shows them to be extremely important.

### WHEN TO USE

Using a design is a dynamic process that has many facets concerning the physical, perceptual and cognitive aspects; how information is absorbed from the design, how it is interpreted and what actions follow as a result are all open to question. Interviews are an excellent method for gathering such information. Interviews are good for revealing participant opinions and subjective judgements, and are useful for gathering insights into complex cause and effect relations.

### PARTICIPANT INVOLVEMENT

The interviewer should aim at giving the interview the flavour of a conversation rather than an inquisition. Achieving this is largely a question of professional expertise and the ease with which the interviewer can establish a rapport with the participant(s). Research projects commonly involve participants in unfamiliar circumstances, and however well they are briefed, participants tend to be uncertain about what to expect. The way an interview is handled is important if the co-operation and interest of the participant is to be obtained and maintained. One way to achieve this is to use *props* as part of the interview. *Props* can be the catalyst that provokes responses and helps to engage participants in a dialogue. Some representation of the design(s) of interest is often used in this way. The presence of a design can be an invaluable aid to explaining points.

### INTERVIEW FORMATS

Interviews may be one-on-one or involve several respondents and several design researchers simultaneously. One of the main stimulants for the evolution of these approaches has been the desire to



explore more productive formats for obtaining information from users. The format of one design researcher and two users has been employed, sometimes referred to as *co-discovery* or *dyads*. In one example, it was used to investigate the problems of users installing a small computer (Comstock, 1983). More generally, this format has been advocated as an aid to protocol analysis and in expert knowledge elicitation (see Chapter 7) where users are required to verbalise their interpretation of a product as a trial progresses (Rubinstein and Hersh, 1984). This has become a widely used approach within user interface evaluation. In recent years, group interviews, often referred to as focus groups, have become much more common in design development work. Focus groups involve one or two design researchers and several participants. The value often claimed for this approach is that participants are stimulated by the observations of their peers, and additional insights emerge which would not otherwise be the case. On the other hand, concerns are that particular individuals, or issues, can dominate a discussion with the effect that important issues are obscured or lost. Good facilitation can counter these tendencies – see Langford and McDonagh (2003).

## DESIGNER AND DEVELOPER INVOLVEMENT

Interviews can be excellent formats for enabling design and developer team members to experience first-hand participant opinions. But an important caution is not to overwhelm/outnumber participants with several interviewers.

## HOW TO USE

### Process Involved

At an early stage, the design researcher needs to decide on the type of interview that an investigation requires. Interviews can be either structured (with predetermined questions and topics), semi-structured, allowing for the respondent to focus more on topics that interest them or being almost entirely open (perhaps using a topic guide to help focus the interview). Whether an open style or closed style of interview is used, the design researcher needs to be clear about the information that the interview needs to deliver. Correspondingly, the design researcher needs to consider question to be addressed, their sequence and the way in which responses will be recorded. The design researcher also needs to consider whether any particular devices, artefacts or other *props* will be used to help the participant(s) answer.

### Expertise Needed

The interviewer needs to be able to put participants at ease and cultivate a rapport with them quickly. In multi-participant formats, the interviewer also needs to be able to achieve a balanced discussion between participants.

## OUTCOMES

Sometimes just talking is at times an inadequate form of expression. So consider the use of different media for recording the outcome of the interview such as writing, drawing and videoed explanations. Whichever approach is used, the purpose of the design researcher is to gain insight into the opinions of the respondent(s) and generate actionable data for the design team.

## EXAMPLES OF USE

Durrant et al. (2009) report on an interview-based study, which explored the use of domestic photo displays in family representation to help critically inform the design of next-generation domestic photo-technologies.

Belloti and Smith (2000) report on a study into the use of an office information management system that featured on-site interviews as a key part of the study.

O'Brien (1982) used group interviews in the context of *participatory design* exercises for the identification of user requirements and subsequent evaluations using product simulations for the design of control rooms.

## QUESTIONNAIRES

### OVERVIEW

Questionnaires are either paper-based or digitally delivered tools that incorporate a series of questions to be answered in a predefined order. Their main purpose is to collect people's responses to questions of fact or opinion in a systematic way so that the data can be easily analysed. They can vary greatly in complexity. In recent years, online presentation techniques have become more common ways to administer questionnaires. Chapter 4 provides detail on questionnaire structure and wording.

### WHEN TO USE

Questionnaires are suitable for any type of question that can be answered *on paper* or *on screen* – either by a verbal or written response. The types of questions asked are either open ended or closed. Open-ended questions do not limit the respondent to a predefined type of answer. Open-ended questions have the advantage that they are likely to generate a wide variety of responses and provide more detailed information but they require post hoc coding during analysis. For large surveys, this type of question can be very revealing but time-consuming to analyse. In closed questions, the respondent chooses between a set of predefined categories. This is an efficient way of collecting basic factual information, and collecting, rating and ranking data for scaling preferences and attitudes. In general, they are very easy to set up and administer providing they are well designed (Oppenheim, 1999; Rea and Parker, 1997 and Chapter 4 of this book). Questionnaires are not good for getting participants to describe complex cause-and-effect relations.

### PARTICIPANT INVOLVEMENT

Questionnaires can be completed either by the respondent or by the design researcher. Often, a combination is used. In either case, questionnaires often form part of a structured interview. Questionnaires may also be used in surveys where participants complete the questionnaire without direct supervision of the design researcher.

### DESIGNER AND DEVELOPER INVOLVEMENT

Designers or developers tend not to participate in administering questionnaires unless they form part of an interview. But it is often useful to involve them in deciding on the topics that need to be addressed and the kind of information that the questionnaire needs to deliver.

### HOW TO USE

#### Process Involved

The key issue for the design researcher is to decide on the questions that need to be asked and the way they should be answered. The design researcher also needs to decide *how* to administer (paper based, phone based, screen based or online), in what context will participants be completing the questionnaire and will it be administered by the design researcher or will it be self-completion.

It is always good practice to draft up the questionnaire and run a pilot. This should include analysing some hypothetical data as a means of checking whether the questions are likely to reveal the information that the design team needs. It is also important to check the layout of the questionnaire, especially if it is to be completed by the respondents at some remote location.

### **Expertise Needed**

Basic questionnaires are relatively straightforward to design, but if the array of questions is large and large quantities of data will be generated, then a specialist in questionnaire design may be required. Questionnaires usually benefit from carefully designing the graphic layout. Experienced graphic designers can be a great help. Finally, the way the data are to be analysed needs to be considered and planned by someone experienced in the statistical analysis of data.

### **OUTCOMES**

Questionnaires can be used to generate various types of data, such as factual statements, ratings and rankings. This can then be used to

- Provide formative or summative evaluations of entire designs or specific sub-components
- Assess usability metrics for a design
- Compare and rank different designs and designed features
- Compile lists of priorities for design fixes and improvements
- Provide detailed user opinions and preferences about a design
- Cross-tabulate and compare user preferences for and opinions about a design by different user groups

## **PROJECTIVE TECHNIQUES**

### **OVERVIEW**

Projective techniques are methods that invite people to express ideas, thoughts and feelings in forms that rely less upon verbal expression and more upon making things, creating or reacting to imagery. Imaged-based methods – such as collage-making, sketching and Lego Serious Play (Cantoni et al., 2009) – offer a level of enquiry that taps people's non-verbal and often implicit cognitive and emotional experience.

### **WHEN TO USE**

Projective methods are suitable primarily for exploring emotional aspects of design or a domain and the meaning that people attach to it. The methods are valuable in situations where people may find it difficult to articulate or reveal attitudes and thought processes verbally. This might include occasions when the topic is inherently non-verbal (e.g. describing spatial or cognitive perceptions), there are social taboos involved (e.g. discussing finance or personal hygiene) and the topic of interest is an abstract or complex one (e.g. entertainment or parenting).

Projective techniques are useful primarily at the exploration phase as a way of uncovering latent feelings and needs and in revealing key qualities that can later be interpreted as early design ideas.

### **PARTICIPANT INVOLVEMENT**

These methods challenge participants to do things that may be unfamiliar to them or that they are nervous about. Participants need to be put at ease, invited to engage in a spirit of fun and reassured that the exercise is about a process of expressing their ideas and perceptions not to produce art.

The role of the design researcher here is to ensure a non-threatening environment and convey supportive curiosity about what and why the participant has produced what they have. Participants themselves are encouraged to explain the meanings and associations of the images. Projective techniques can be applied one-on-one with participants or in group sessions.

### **Designer and Developer Involvement**

In group sessions, it can be useful to engage design team members, even clients, in the creative process. Workshop settings provide more scope for their involvement in the process to encourage personal insights and also a sense of shared discovery. When conducted one-on-one, design team members may be remote from the creative event itself. In reporting discoveries, there is great value in exposing selected examples of the actual products themselves – collages, maps or models with commentary to provide interpretation – to illustrate specific points of a more general nature. Of collages, for example, Serpiello (2001) writes: 'A project report – text upon paper, is easily ignored and hardly passionate, but a three-foot poster plastered with images of what is most important to the client's customer packs a punch'.

### **How to Use**

#### **Process Involved**

Projective methods can be used in association with one-on-one interviews or in activity/workshop settings. The methods all need to be introduced to participants after they have become comfortable and feel safe and reassured that you are genuinely curious about their internal mental world and reactions. This would usually be at least 20 min into an open-ended interview session or after group warm-up activities at a workshop. All these activities can be done in a wide-open exploratory manner, but it is helpful to provide participants with a clear request and a fairly short time limit so that their responses are spontaneous rather than rationalised, which would defeat the point. It is best to take the lead in deciding how much discussion goes on during their creation phase. Some people like to talk aloud, while they create and enjoy explaining and being asked questions. Others like to focus exclusively upon what they are making and engage in conversation when they have finished. In either case, the participant is invited to explain the elements and layout of their creation, why they selected specific images, arranged elements as they did or chose a particular shape. This enquiry requires sensitivity. Some decisions will have been made at a less conscious level than others (and these may be the most interesting) so it is important not to force the participant to make up explanations. The discussion should rather take the form of a mutual discovery process where the design researcher is the encouraging audience. Participants need to be provided with whatever tools they need. Examples are clay, foam board, paper or card, drawing instruments, stickers, images culled from magazines or photocopied from books, photographs, word labels on cards, glue, scissors and modelling materials.

#### **Expertise Needed**

People with excellent social skills who are able to put participants at ease, provide clear instructions and perhaps of most importance are skilled interviewers. The value of these methods often lies in the participants' explanations of the significance of what they have created. It can also, in group sessions, be very helpful to have skilled model makers and illustrators on hand to help participants express their ideas.

### **OUTCOMES**

The methods provide an alternative form of expression for thoughts and feelings. The process itself enables participants to unearth associations and make discoveries about their own perceptions and use the design as a starting point for discussion and personal story telling. The methods provide insight into the role and meaning of personal and culturally relevant rituals; mental models and

cognitive structures; emotional valence of specific activities and design elements; and significance, associations and cause and effect relating to experience. Thought must be given, though, to how the results of the sessions will be communicated to others on the design team; video and/or photo capture of output is often critical for communicating findings.

## EXAMPLES OF USE

Lynch (1960) shows examples of people's maps and drawings to explore how they understand and perceive local and world geography.

Serpiello (2001) describes how collage-making was used in a design programme involving the development of new kinds of water-delivery mechanisms and plumbing fixtures.

Cantoni et al. (2009) demonstrate how the Lego Serious Play techniques can be adapted to allow participants to use Lego to model future services.

## ROLE-PLAYING

### OVERVIEW

Role-playing methods include a range of immersive techniques in which members of the design/development team are asked to personally take part in using products, environments, events or services from the perspective of a user. The methods involve the design team, sometimes with props to assist them (see *empathy tools* later), taking on a specific set of characteristics – motivations, abilities, limitations – and performing or participating in activities in the real-world or contrived environments. Role-playing is used mainly by design teams alone, but it is an approach that also provides an opportunity for participants and a design team to work together to explore a particular design issue.

### WHEN TO USE

Role-playing is valuable at all stages of the design process. In the early exploratory phase, people can learn about design requirements to support various users and roles. It is beneficial as a direct method of generating ideas in situ as role-players are engaged in life-like activities and contexts. As prototypes and simulations are developed, role-playing enables the players to evaluate their strengths and weaknesses. And finally, once a design is implemented, role-playing provides an opportunity to gain insight into the kinds of experiences that people have in using and reacting to the design.

### PARTICIPANT INVOLVEMENT

Participant involvement in role-playing techniques occurs in two ways. First, participants can directly act out the roles they would have in reality, effectively demonstrating scenarios they would normally be engaged in, as users. Second, participants can act as expert advocates helping the design team to successfully (and more accurately) partake of the users' experience. This latter role can be particularly important when users have specific accessibility needs that it might otherwise be hard for the design team to role-play and understand, without guidance.

### DESIGNER AND DEVELOPER INVOLVEMENT

Role-playing methods allow you to engage the design team in first-hand explorations and discoveries about the design domain from a range of perspectives different from their own. It is useful when you want team members to deeply understand the issues facing different users in different contexts.

This personal understanding is valuable as inspiration in generating, evaluating and refining design ideas to take diverse users' issues into account. In contrived settings particularly, when it is not their turn to *play*, other team members may be available to observe the action of their colleagues. This provides an opportunity for another source of insight about issues and solutions.

## How to Use

### Process Involved

Role-playing methods require a balance of sufficient structure and direction for team members to feel comfortable participating, and flexibility for them to improvise and make their own discoveries through the process. You need to define the role you are asking the players to play, and provide some basic props that will help with that. For example, each player can be given a reference card that describes the personal characteristics of the abilities, motivations, goals and specific activities for their role. These require careful planning to ensure that you are exploring an appropriately diverse set of conditions. Simple props, modelling materials or prototypes may be used to enable particular contexts for the action to be created. The players may also be provided with contrivances designed to limit specific abilities so as to simulate impairments, for example earplugs to limit hearing, latex gloves to reduce sensation and bandaged joints to restrict movement. Devices such as these are often referred to as *empathy tools* and specifically help designers to begin to experience the world similarly to the users.

During or immediately following enactment, it is important to facilitate reflection about individual experiences and what's been learned. This could be by asking players to keep diaries, for example, or to participate in a group session where individuals and observers share their discoveries in a semi-structured way, for example first listing problems and issues and then brainstorming solutions. Role-playing methods allow a design team to engage in first-hand explorations and discoveries about the design domain from a range of perspectives different from their own and to understand deeply the issues facing different users in different contexts. This personal understanding is valuable as inspiration in generating, evaluating and refining design ideas to take diverse users' issues into account.

It is important to recognise that role-playing, while it has the beneficial aspects of creating personal insights into the experience, needs and perceptions of other people, is to a large extent based upon imagination and fiction. It should never be used as a substitute for information gathering from the real world and attempts to understand the nature of other people's reality thorough direct interaction with them (hence the suggestion to use real end users to help structure and guide the role-playing activity).

### Expertise Needed

Running a successful role-playing session benefits from good social and group facilitation skills, but it relies also on a team's willingness to participate. It can help to give team members confidence in overcoming natural shyness to work with skilled teachers of improvisation methods, often actors themselves.

## OUTCOMES

Designers' own experiences while role-playing will engender empathy for users, enabling them to appreciate issues including physical difficulties, cognitive confusion and contextual issues in use. By encouraging members to enact the role of potential users, you can facilitate a common team understanding of the critical user issues. Information becomes more vivid and engaging when it resonates with personal experience. If designers and clients can have informative personal experiences, it is easier for them to grasp the issues and feel greater empathy with both the people who will be affected by their decisions and the experiences users may face.

## EXAMPLES OF USE

Moore (1985) describes the author's experiences while she was disguised as an elderly person and is a classic example of learning through role-playing about the needs and challenges of people different from ourselves.

Buchenau and Fulton Suri (2000) describe several applications of role-playing methods including body-storming.

Anderson and McGonigal (2004) discuss *place storming* and the importance of using spatial context as a prop when conducting role-playing-style sessions, in particular for designing location-based services (such as pervasive games).

Clarkson et al. (2007) introduce their *Inclusive Design Toolkit* and provide a variety of techniques and simulation tools for designers to directly develop empathy for users with accessibility concerns.

## SCENARIO BUILDING

### OVERVIEW

Scenarios mean different things to different people (Campbell, 1992). But the common thread of scenario building, in a design context, is about creating one or more fictional portrayals involving specific characters, events, products and environments, to explore and envision design ideas, technologies or issues in the context of a realistic future. These stories may take many forms: text narrative, annotated sketches, cartoons, photographs, video or live enactment. They may vary in scope and scale, involving entire processes or events over a period of time, storyboarding of a product's interaction behaviour or vignettes that portray a brief moment or single event. Scenarios should not be just fantasies, but carefully constructed projections of anticipated usage situations that are based on available evidence.

### WHEN TO USE

Scenario building is a powerful exploration, prototyping and communication tool, particularly useful early on in the design process, before committing substantial resources to detailed design and development. Campbell (1992) identified four distinct purposes for future-focussed scenarios: to illustrate the use of a system, to evaluate system functions, to design attributes or features and, though of less value in a design context, to test theory. In phases 1 and 2, design scenarios are useful for exploration and idea generation, to learn and respond to usability and lifestyle issues that will be relevant to a particular group of users, and how contextual and procedural factors might be accommodated. In phase 3, they are useful as an evaluative and diagnostic tool, to discover faults that exist in a design concept and how they might be rectified or how alternative or competing solutions, interaction protocols might play out in practice.

### PARTICIPANT INVOLVEMENT

Participants in some cases may be directly involved in developing scenarios about their current or future activities. In others, they may themselves not be directly involved but will provide the base material about real people in real contexts which will inform the development of a small set of fictional characters, personas and settings that feature in the stories. The value of abstracting lessons from a fictional character set is threefold: it allows the team to know intimately and talk publicly about them without being invasive of real people's lives, it allows the team to easily grasp and relate to a huge range of important human differences embodied in a memorably small group of characters and it allows the team to project people into future situations.



## DESIGNER AND DEVELOPER INVOLVEMENT

Scenario building may be the responsibility of an individual on a team, often the user advocate, but often it may also involve participation of all members of the design team. Since it benefits from complementary skills of analysis and synthesis, verbal and visual fluency, there are roles for diverse talents. Moreover, one of the main benefits of scenario building is that it provides a shared view of user issues that the entire design and development team can refer to. For this reason, it is valuable to develop a sense of shared ownership through a high level of participation by the design and developer teams in creating scenarios and interpreting the lessons. Scenario building offers the opportunity to involve the developer organisation in appreciating the user perspective on the design in question.

## HOW TO USE

### Process Involved

Scenario building has its roots in the more traditional techniques of user profiling, task analysis and system ergonomics (see Chapters 6 and 9) but is increasingly becoming influenced by other disciplines such as film studies and narratology. The process starts by analysing the relevant human characteristics, motivations, tasks, social and technology trends and contextual issues that need to be considered. Elements that can be represented in scenarios include (1) a set of individual users, detailed with respect to abilities, lifestyle and circumstances; (2) issues, goals, tasks and situations; and (3) the design itself. The design might be a well-defined proposition or a loosely defined idea to be developed through the process of building a scenario. Next, the elements are woven together to create a coherent, believable, unbelievable or even provocative story or vignette. Like all creative endeavours, scenario building is itself an iterative process. Gaps may appear in knowledge about the people, contexts or technology, which prompt a further round of analysis and restructuring. Some proponents of specific scenario building techniques, which use video prototyping, where future products are imagined on screen, have suggested the use of narrative tools such as never showing the designed artefact being discussed by characters in a scenario (see Briggs et al., 2012) or the use of multiple versions, to demonstrate negative and positive scenarios around the same concept (see Mancini et al., 2010). These techniques are used to primarily enhance the discursive potential of the scenario developed.

### Expertise Needed

The key skill in scenario building is the ability to synthesise and weave together diverse information concerning people, their characteristics and motivations with contextual issues and technology, product and service ideas to create a believable or impactful story. Narrative development skills are essential, and depending upon the medium of expression, technical skills in video, illustration or photography can be useful. Frequently, it is beneficial to have an expert in the appropriate technologies available.

## OUTCOMES

Scenario building has value both as process and product. As a process, creating scenarios has the effect of forcing the team to think through usage issues early. With very low investment, it provides an easy way for users, team members and clients to explore, discuss and grasp how a design might work in practice. It encourages the formulation and exploration of multiple questions about usage and context over time as questions arise such as ‘what if... he doesn’t notice the blinking light? ... someone interrupts him now? ... he has brought a child?’

As a product, once built, scenarios promote understanding and identification amongst a design team and clients with the users portrayed as specific individuals. They provide a succinct way of representing and realising the temporal and spatial consequences of design decisions as they may

play out in reality and so allow qualitative evaluation of dynamic aspects of interactions with a design. Scenarios have the effect of bringing to life a detailed task analysis and embedding it in context so that it becomes an easily shared portrayal of positive and negative impacts of context upon interaction with a design. Its value is in providing easily shared evidence of human factors issues for design teams. A further advantage is that the resulting scenarios provide stimulus material for use in evaluation by users, in interviews or in workshop or focus group settings.

Fulton Suri and Marsh (1999) discuss some dangers to be avoided in scenario creation. Since they are inherently fictional, there may be temptations to avoid or gloss over difficult situations, rely on stereotyped descriptions of people or offer only single solutions and justify weak ideas.

## EXAMPLES OF USE

Carroll et al. (1997) report on the development of requirements for a virtual physics laboratory based on creating scenarios together with staff and students.

Fulton Suri and Marsh (1999) give examples of scenarios used in various consumer product design projects.

Rosson and Carroll (2002) provide an extensive discussion of the use of scenarios in the development of computer systems.

Sterling (2009) discusses the potential of *design fictions* in scenario explorations for next-generation technologies.

Blythe and Wright (2006) introduce the notion the *Pastiche scenario* extolling the virtues of caricaturing potential design features/scenarios to draw out associated user values.

Lang et al. (2013) used scenarios with adolescent participants to inform understanding of requirements for the design of medical devices.

## USER TRIALS

### OVERVIEW

A user trial is primarily about creating an environment, often in a usability laboratory, that enables the interaction between a design and a user to be systematically examined and measured under controlled conditions (see Chapter 2). User trials enable a design researcher to measure the effectiveness of designs from both quantitative and qualitative points of view. User trials involve applying the basic principles of experimental design, measurement techniques and data analysis that apply to formal *experiments*. So, to adapt the description Chapman (1959) gave for the experimental method, a user trial is ‘a series of controlled observations undertaken in an artificial situation with the deliberate manipulation of some variables in order to answer one or more specific questions about the effectiveness of the design’. In practice, when user trials are used as part of a design development programme, they tend to examine the initial learning phase and not habituated use.

### WHEN TO USE

A user trial is the classic method to use when a design team needs to evaluate a design (or a specific feature of a design) under controlled conditions. The design is usually in the form of a simulation or prototype, but the method lends itself equally well to the evaluation of implemented designs in actual use. A user trial is an excellent method when a design representation can be evaluated using usability (and similar) metrics. Metrics may include specific formal or informal performance standards that the design should meet. So the method can be used in phases 2, 3 and 4. In phase 2, the method can be used to investigate user needs and requirements. For investigations of this sort, the design researcher needs to be aware of different types of objectives and adapt accordingly. More traditionally, the method is often used in phase 3 from very early stages of concept development

up to the evaluation of complete prototypes. The method can be adapted to suit design evaluation under real-world conditions. In this context, the design researcher needs to be aware that attempting to evaluate a design under *controlled conditions* in real-life circumstances can create artificial situations. User trials can be adapted for the following:

- *Exploratory studies*: What usability issues are relevant to a particular group of users? What expectations do users have of early design concepts?
- *Diagnostic studies*: What faults exist in the design, and how might they be rectified? Design faults can include operating procedures, physical configurations, instructions for use and labelling.
- *Measuring performance*: Does a particular design meet specific performance requirements? They can be used for benchmarking a design, both *within* and *between*: within itself as design versions evolve to check on progress towards the design objectives, and between the design and competing solutions. Competing solutions may be comparable designs from other manufacturers or other designs from within the developer's own organisation.
- *Investigating contextual factors*: How does the design interrelate with the context in which it will be used?

## **PARTICIPANT INVOLVEMENT**

User trials are excellent for demonstrating user perceptions and experiences of a design to the design and developer teams. User trials can use different formats in terms of how to involve participants. Two or more participants can be involved in one user trial. Examples include the *co-discovery technique* (Kemp and Gelderen, 1996). Monk et al. (1993) have also developed a *co-operative evaluation* technique specifically aimed at evaluating early design proposals.

## **DESIGNER AND DEVELOPER INVOLVEMENT**

A trial enables the design team, developer teams and clients to remotely observe the trials in real time. Real-time or post hoc review of videos can be extremely persuasive when design review discussions get underway. There is potential for a high level of participation by the design and developer teams in observing, analysing and interpreting the outcomes. There is also great potential for involvement by others in the developer organisation in appreciating the user perspective on the design in question.

## **HOW TO USE**

### **Process Involved**

Probably, the three most important issues to consider when setting up a user trial are the key aspects about the design that need investigating, the form in which the design will be represented and the location of the user trial. The design researcher then needs to consider the tasks the participants will carry out, the performance criteria that will be used and their associated data collection method(s). The tasks need to follow a pattern that makes sense to the participants in terms of how the design would be used in practice. The tasks should be also designed so that the investigation can get at the aspect(s) of the design that concern the design team. How the data are to be collected will also be heavily influenced by the choice of location. Often user trials incorporate interviews and questionnaires. It is very common to run user trials in a usability laboratory in which audio and video data can be recorded.

### **Expertise Needed**

Running user trials requires expertise in experimental design, statistical analysis of data and interview techniques. Expertise is also required in audio and video data analyses. In terms of the user trial itself, the design researcher also needs staff skilled in creating a rapport with participants.

## OUTCOMES

User trials typically generate quantitative and qualitative data, and often in large amounts. The specific form of the data obviously depends on the metrics selected. Typical examples are task times, error scores, physical fit, comfort scores, attitudinal measures, general impressions, etc.

## EXAMPLES OF USE

Fulton Suri (1993) reports on the use of user trials to evaluate information graphics.

Kirk and Stanton Fraser (2006) report on the use of user trials to evaluate different communication technologies in support of remote collaboration in physical tasks.

Lavender et al. (2014) describe a user trial experimentally comparing three forms of hand-carried stair descent devices for evacuating mobility-impaired occupants from high-rise buildings.

## USER WORKSHOPS

### OVERVIEW

User workshops are events where a group of participants work on a design issue facilitated by the design researcher. A critical characteristic is that such workshops generate tangible evidence, such as drawings or models, for example, of the participants' perspective on design(s) for a particular purpose. In this respect, they differ significantly from focus groups which typically rely upon discussion only. The design issue chosen for user workshops needs to be one which participants have the knowledge and expertise to tackle. User workshops can be used to explore such issues as functional needs, workplace layout and general configuration of a user interface. The location for user workshops can be various. It is a question of what suits the purpose of the investigation and what is convenient. In general, user workshops are best supported by flexible informal spaces in which participants feel free to move about, make things (if needed) and change things, without fear of *making a mess*. The great value of such workshops is that participants often get to express their ideas through media other than simply verbal or written. While participants may be working on particular details of a design, they can be also stimulated to address more fundamental concerns in terms of *needs and requirements*, characteristics that make a design acceptable or not, and so on. It is this kind of output that is what makes user workshops so valuable for design teams.

### WHEN TO USE

The idea of user workshops grew out of the simple idea that participants and designers could work productively together to resolve design questions. The development of the concept has evolved to encompass a wide variety of specific approaches that cover all phases of the design process. In phase 1, they can be a useful complement to ethnographic methods as a way of participants elaborating on perceived needs and requirements, and contextual issues. In phase 2, they can support the definition of the design space in terms of requirements, contextual issues and the features of a design that will make it suitable for the purpose intended. In phase 3, they can be used to support the definition and evaluation of specific design proposals or, in the case of phase 4, fully implemented designs.

### PARTICIPANT INVOLVEMENT

Participant involvement revolves around particular activities set out by the design researcher. The activities in themselves usually produce particular tangible outcomes. But one of the key objectives of any user workshop is typically to get the participants to interact with each other (as well as designers and/or developers if they participate) and thereby benefit from the cross-fertilisation of their individual experiences.

## DESIGNER AND DEVELOPER INVOLVEMENT

If designers and/or developers also take part, they can often collaborate with participants where the designers facilitate the articulation of participant ideas. This can be an excellent way to foster interaction between the design team and participants. The result is often that the designers not only get to understand particular ideas but also to develop a more in-depth appreciation of the people who are expected to use their design.

## How to Use

### Process Involved

As with all forms of workshops, the design researcher needs to work out a clear activity agenda that successfully facilitates the achievement of the chosen objectives. All the participants need clarity about the type of outcome expected from them. The activities and the outcome expected need to be carefully selected so that the participants do not have to climb a steep and difficult learning curve. The participants should be able to become productive quickly and thereby focus on the subject of the workshop and not be hindered by a lack of skill or expertise. If members of the design team are involved, it is important that the design researcher takes care to brief them beforehand on the role that they should play.

### Expertise Needed

However carefully the agenda is designed, workshops require an excellent facilitator to ensure that the objectives of the workshop are achieved.

## OUTCOMES

The great value of such workshops is that participants get to express their ideas through media other than simply verbal or written. This can include representations of how a design might appear, visual descriptions of how a system should behave, video walkthroughs showing how a user might interact with a system and life-size models of workplace layouts and hardware configurations. While participants may be working on particular details of a design, they can also be stimulated to address more fundamental concerns in terms of *needs and requirements*, characteristics that make a design acceptable or not, and so on. User workshops can be excellent for helping a design team to identify critical design features and rank design issues in importance.

## EXAMPLES OF USE

Eggen et al. (2003) report on a study into the home in which user workshops featured as a method for developing new application concepts.

Mackay et al. (2000) report on a study into a GUI design for a *Petri Net* application using participatory design techniques and video-based tools.

Vines et al. (2012) discuss the use of participatory design workshops for eliciting feedback on new concepts for banking technologies for the over-80s.

In this volume, Gyi et al. (Chapter 34) present a range of participatory ergonomics methods.

## FUTURE DIRECTIONS

In this chapter, we have introduced a broad range of different ways of involving people in design. We now discuss some future directions and focus on a number of issues that will influence the development of methods and the professional practice of designing.

## ORGANISATIONAL AND TECHNICAL BARRIERS

As designs become more complex, technology more powerful, commercial pressures more severe and resources more expensive, there is greater need to tackle the human impact of designs. Developers, designers and providers of products, services, environments and media are increasingly interested in methods that enable them to better anticipate people's needs and desires so that they can provide more successful designs in both commercial and human terms. Implicit in this interest is the drive to base design decisions as far as possible upon *evidence* rather than on *opinions*. This is precisely where the effective use of methods, of the kinds we have described here, can play a vital part. But the relevance of such *methods* to mainstream commercial design development will depend upon their effective adoption and use. They need to demonstrate value in creating and inspiring successful design directions and in efficiently guiding implementation.

## DESIGN AS A STRATEGIC BUSINESS ADVANTAGE

Many companies in many business sectors have long valued design as a central function in their business that complements marketing, manufacturing or engineering. In recent years, design has been recognised by more companies – beyond its obvious function in generating products – as an important way that businesses can favourably differentiate themselves from their competitors – see Peters (1998, 2000) for a discussion of design from a business perspective. With this key business role, it is even more important that designers successfully anticipate and provide solutions that elicit positive responses from people. Involving people in the design process has much to offer in the form of a range of ways to learn about people's behaviour and motivations (see Buxton, 2007 for extended discussion of this).

In the longer term, implementation of HCD processes in industry will depend on managers recognising that ensuring user satisfaction with their designs needs to be embedded in the way their organisation manages product development and product implementation. This requires a more systematic approach to design quality and usage issues than is typically the case. In other words, design quality and usage issues need to be incorporated into the specific design objectives that product developers have to meet. On this basis, it is then possible to select the appropriate methods and maximise their benefit by focusing on the design issues in question. The principles are easily stated but, in practice, are very difficult to implement. Gould (1988) and Vredenburg et al. (2001) present valuable discussions on this topic (see also Grudin, 1990).

One of the major benefits of using the methods outlined in this chapter is that they help to make usage issues tangible and manageable. The usage issues get expressed in terms and in ways that allow design teams to manage in a balanced way the inevitable compromises and trade-offs with the technical, commercial and project planning issues. The methods also confront design teams with the *real questions* concerning the purpose of a design and the value it has for people's lives, in contrast to just checking on device performance as an artefact in itself.

## PROVING THE VALUE AND RELEVANCE OF METHODS

There is considerable scope for research work aimed at assessing the most effective methods for ensuring that human issues are taken into account within design development. In terms of the more traditional human factors and usability methods, Meister, as long ago as 1986, noted the lack of attention given to evaluating methods that are used to guide design and called for further efforts to improve our understanding of their value – see also Anderson and Olson (1985) and Karat et al. (1992). It is a critique that is still relevant today. Many of the technical issues discussed are of long term and academic interest lying beyond the pragmatic concerns of most design researchers within industry. Design researchers within an industrial design–development environment, of necessity, have to take a short- to medium-term view. Nevertheless, it is of great practical importance to designers and developers to review critically where and when specific methods are most



effective, reliable and valid. This kind of critical appraisal will undoubtedly be a focus for the future. However, the real and changing world of designs-in-use – where an enormous number of variables determine a design's success – will be a very challenging arena in which to investigate the effectiveness of methods with high levels of confidence. The successful implementation of methods is likely to depend on demonstrating their effectiveness in terms of the following:

- How far is it possible to predict how a design will perform in practice?
- Can the performance of designs in use be improved by evaluating their performance during development?
- The cost–benefit ratio of development to organisations of improving the quality of designs, ensuring that people's needs are met and that design deficiencies do not reach the market.
- Reducing the time it takes for investigations to reveal valuable insights to the design team. This means reducing the time it takes to execute investigations, to interpret the results and to incorporate the results into design work.
- How people's involvement in the design process positively changes people's perceptions of the organisation and the designs it sells.

### **EARLIER INVOLVEMENT AND FASTER RESPONSE TIMES**

As mentioned earlier, a key issue is the demand for earlier and more timely input concerning human issues, to ensure that efforts are directed appropriately at all stages in the process. Tools and methods are required which reduce the time taken to carry out investigations, including set-up, collection, interpretation, reporting and application of discoveries. New technologies – for example digital audio and image capture, networked communications, the Internet – not only offer the design researcher many opportunities to gather and disseminate a wealth of data rapidly but also bring with them the potential handicap of gathering much more data than are useful or necessary and adding time for collation and analysis. In many cases, alert observation and thoughtful use of traditional tools – paper and glue, face-to-face communication – still prove to be very effective. But there is clearly room for improvement in tools and methods now available to help reduce turnaround times and speed up the information transfer process.

As HCD continues to gain ground both in organisations that develop designs and in the professional practice of designing, the scope and boundaries of HCD concerns will evolve. And as these changes occur, the methods at our disposal must also evolve. We do not have the space to discuss the issues in detail, but nevertheless we want to highlight some major interrelated trends that we expect to have a significant influence on the development of our methods in the coming decades.

### **ECOLOGICAL VALIDITY**

How well do the methods used help design teams accurately anticipate the way designs will be used in practice? In the final analysis, the value of HCD is in the extent to which organisations can increase their rate of success both in introducing products and services that positively delight their users and customers and in reducing the risk of design failures and errors. Errors in design may have no more serious consequences than mild inconvenience. But on the other hand design errors may at times have lethal consequences for a few or, worse still, for many people. At the heart of our methods lies an explicit or implicit usage model, and it is the validity of the usage model on which rests the value of HCD. The methods we use are either concerned about understanding and modelling use, or assume a certain model of use to either guide design decisions or evaluate the suitability of solutions. So a major issue for the future is the accuracy and reliability of the usage models design teams use. The scope and level of detail to which any usage model must go obviously depends on the type of design required and the organisational context of the design team. There are clearly, in principle, many factors to consider. In practice, a design team has to take a pragmatic approach



and make its own best estimate of the usage model required, the critical factors to consider and the extent to which they can rely on assumptions rather than explicit evidence.

## DESIGN CRITERIA

HCD as a concept relies on the use of explicit criteria both to guide design decisions and to evaluate the outcomes. These criteria should reflect the concerns and interests of the people who will use the design. In recent years, an important and growing debate has emerged about how methods will need to adapt to cope with the variety of human issues that affect the success of a design. Design practitioners of all professional persuasions have begun to look for methods and frameworks that embrace the greater variety of human issues. The challenge is to discover what is ‘simultaneously useful (needed), usable (understandable) and desirable (wanted)’ (Sanders, 2002). This is particularly true in the realm of designs for consumers, where emotional experiential aspects relating to appeal, aesthetics, image and lifestyle-fit are crucial to a product’s success. A central theme in this debate is the way people *experience the design* (see, e.g. McCarthy and Wright’s 2004 *Technology as Experience*). Correspondingly, the debate also centres on the ways and means by which the *experience* can be both anticipated when designing and measured when solutions are evaluated. Affective, social and cultural issues all contribute to and influence people’s experience of designs but have traditionally been neglected by the human factors profession (Jordan, 2000). But other professional groups concerned with psychology, market research, sociology and anthropology (e.g. Csikszentmihalyi, 1991) have made attempts to address these issues and discover tools and methods that enable these issues to be managed. In the recently published ‘Being human’ report (Harper et al., 2008), exactly these issues of how more humanistic concerns can be introduced to the technology development process during design phases have been explored. In a similar vein, there is increasing interest in the affective aspects of designs, their aesthetical impact on the users’ emotions and the significance of the design in terms of the social context of use. The central question is which are the critical criteria that design teams need to design to, and do we have the corresponding methods that enable the criteria to be used effectively to guide design decisions and evaluate solutions?

## AESTHETICS AND PREFERENCES

Even today, there is little established theory regarding the human aesthetic response, although practitioners are beginning to embrace the issue, particularly in relation to product design, for example Desmet and Hekkert (2002), Fulton Suri (2002), Jordan (2000), McCarthy and Wright (2004). Crozier (1994) provides an overview of contributions to questions of aesthetics and preference including Gustav Fechner’s experimental psychophysics, explorations of the *golden section*, the Gestalt theories of human perception, Berlyne’s theory linking aesthetic preferences with physiological arousal and Arnheim (1974) making connections between art and psychology, and with McCarthy and Wright (2004) drawing on the philosophical foundations of Dewey’s Aesthetics. None of these theoretical approaches is able to fully explain or predict the way people will react to designs, not least because aesthetic preferences represent only a part of the significance of objects for people (Desmet and Hekkert, 2002). However, our tools and methods must continue to address these matters even when we lack predictive theory and principles.

## PERSONAL AND SOCIOCULTURAL

Social and cultural factors are intrinsic to some kinds of design, such as shared spaces, communication and collaborative systems. But there are also subtle and complex social and cultural implications for all kinds of designs. Designers, marketers, anthropologists and consumers know, as Helga Dittmar (1992) says, ‘... material possessions have a profound symbolic significance for their owners, as well as for other people ... they influence the ways in which we think about ourselves and

about others'. Miller (1997) suggests that we need a more formal discipline of *social ergonomics* and notes several ways in which designs have social significance for people:

- As tools of non-verbal communication, Goffman (1971)
- Establishing meaning about ourselves and our lives, Csikszentmihalyi (1991)
- Stimulating certain kinds of social behaviour – for example mobile phones demand a response wherever we are, or the way we talk too loudly when listening to a personal stereo
- Affecting our agency – for example when the computer is down or busy, we cannot communicate with our family
- Mediating social interactions – for example gathering places like doors and thresholds can be complicated by automatic doors and door closures
- Ascribing personality to designs – for example when a device or service seems *stupid*, *friendly* or *demanding*

These considerations will continue to be important aspects to be addressed by HCD methods.

## SUSTAINABILITY FRONTIER

Since its origin in a concern with fitting tasks and environments to human need, ergonomics in design has expanded its scope. Indeed, the notion of *human fit* has evolved over time from physical and perceptual to cognitive to social to cultural to ecological fit. We are right now at the brink of this ecological frontier. Sustainability issues loom in the face of the human creative and commercial enterprise of design and production of products, services and environments. Customers and users, especially in the leading industrially developed nations, are becoming increasingly aware of the ways that our individual behaviours and the products and services we consume are negatively impacting quality of life in a holistic sense. Businesses and consumers are beginning to grapple with systemic issues relating to human use of the earth's resources and subsequent destructive interference with natural cycles of regeneration and replenishment. If the underlying motivation for design is to enhance human existence and our *quality of life* collectively and individually, then our design practice needs to encompass consideration of the following:

- Ways to design products, systems and services that support and encourage ecologically responsible behaviour. (What correct/incorrect mental models of ecologically responsible behaviour exist in the minds of designers and users? How are these reflected in design? How can desirable behaviours can be encouraged and simplified by design?)
- A regeneration of concern with fundamentals of human health and well-being. (How can design support and encourage health and wealth production for the underprivileged? How can design ensure clean air and water as a widespread biological and economic necessity?)
- Establishing greater understanding of cultures and value systems other than our own. (What methods are appropriate and useful in working with people from different educational, economic and cultural experience from our own and in learning about their needs, desires and perceptions?)

For extended discussion of the importance of sustainability to interactive technology design, see Blevins (2007).

## CREATIVE AND INTEGRATED METHODS

The overview by Stanton and Young (1998) of ergonomics methods used in product design reveals a clear emphasis on evaluative and analytical tools, whether quantitative or qualitative in nature. No longer is it appropriate simply to evaluate what others design and produce, or to work as isolated

purist specialists on the sidelines of design. Designers and developers have been frustrated sometimes by the results of E/HF analyses and evaluative studies. Such methods provide information about people's capabilities, problems that arise and people's reactions to specific design variables. But as we have seen, people's response to design is more complex than the traditional scope of this work, and information does not, by itself, lead to design solutions. Design requires synthesis and creativity – integration of human and technology capabilities to create coherent and workable designs. To make useful contributions, we need to emphasise and explore methods that bridge this gap between analysis and synthesis, to help translate human factors information into a form that stimulates well-conceived HCD ideas (Hasdogan, 1996). Our energies must be directed not simply to providing information but in engendering a human-centred and people-inspired approach to tackling design issues. Many of the methods explored in this chapter are designed to do just this, through activity, reflection and imagination by participants, designers and developers together exploring and creating designs that will be beneficial for people. We hope that this will inspire the development of more new and effective methods, and that their use will become more widespread and well established in HCD.

## REFERENCES

- Adlin, T. and Pruitt, J. (2010). *The Essential Persona Lifecycle: Your Guide to Building and Using Personas*. Morgan Kaufmann. Burlington, MA.
- Anderson, K. and McGonigal, J. (2004). Place storming: Performing new technologies in context. In *Proceedings of NordiCHI 2004*. ACM Press, New York, pp. 85–88.
- Anderson, N.S. and Olson, J.R. (1985). Methods for designing software to fit human needs and capabilities. In *Proceedings of Workshop on Software Human Factors*. Washington, DC: National Academic Press.
- Arnheim, R. (1974). *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA: University of California Press.
- Bainbridge, L. and Sanderson, P. (1995). Verbal protocol analysis. In J. Wilson and N. Corlett (eds.), *Evaluation of Human Work*, 2nd edn. London, U.K.: Taylor & Francis Group.
- Bannon, L. (1991). From human factors to human actors: The role of psychology and human-computer interaction studies in system design. In J. Greenbaum and M. Kyng (eds.), *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, MI: Lawrence Erlbaum Associates.
- Bellotti, V. and Smith, I. (2000). Informing the design of an information management system with iterative field-work. In *Proceedings of Designing Interactive Systems (DIS 2000)*. ACM Press, New York, pp. 227–237.
- Blevins, E. (2007). Sustainable interaction design: Invention & disposal, renewal & reuse. In *Proceedings of CHI 2007*. ACM Press, New York, pp. 503–512.
- Blythe, M., Overbeeke, K., Monk, A. and Wright, P. (eds.). (2003). *Funology: From Usability to Enjoyment*. Dordrecht, Netherlands: Springer.
- Blythe, M. and Wright, P. (2006). Pastiche scenarios: Fiction as a resource for user centred design. *Interacting with Computers* 18(5), 1139–1164.
- Bodker, S. (2006). When second wave HCI meets third wave challenges. In *Proceedings of NordiCHI 2006*. ACM Press, New York, pp. 1–8.
- Boess, S., Saakes, D. and Hummels, C. (2007). When is role playing really experiential?: Case studies. In *Proceedings of TEI 2007*. ACM Press, New York, pp. 279–282.
- Bowles, C. and Box, J. (2010). *Undercover User Experience Design*. Berkeley, CA: New Riders.
- Brabham, D. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence: The International Journal of Research into New Media Technologies* 14(1), 75–90.
- Bradley, J., Benyon, D., Mival, O. and Webb, N. (2010). Wizard of Oz experiments and companion dialogues. In *Proceedings of BCS HCI 2010*. Swindon, U.K.: British Computer Society, pp. 117–123.
- Briggs, P., Blythe, M., Vines, J., Lindsay, S., Dunphy, P., Nicholson, J., Green, D., Kitson, J., Monk, A. and Olivier, P. (2012). Invisible design: Exploring insights and ideas through ambiguous film scenarios. In *Proceedings of DIS 2012*. ACM Press, New York, pp. 534–543.
- Buchanan, M. and Fulton Suri, J. (2000). Experience prototyping. In *Proceedings of DIS 2000*. ACM Press, New York, pp. 424–433.
- Buxton, B. (2007). *Sketching User Experiences*. San Francisco, CA: Morgan Kaufmann.
- Buxton, B., Greenberg, S., Carpendale, S. and Marquardt, N. (2011). *Sketching User Experiences: The Workbook*. Waltham, MA: Morgan Kaufmann.
- Campbell, R.L. (1992). Will the real scenario please stand up? *SIGCHI Bulletin* 24(2), 6–8.

- Cantoni, L., Marchiori, E., Faré, M., Botturi, L. and Bolchini, D. (2009). A systematic methodology to use LEGO bricks in web communication design. In *Proceedings of SIGDOC 2009*. ACM Press, New York, pp. 187–192.
- Cantwell, D. and Stajano, A. (1985). Certification of software usability in IBM Europe. *Ergonomics International 85: Proceedings of the Ninth Congress of the IEA*. IEA, Zurich.
- Carroll, J.M. (ed.). (1995). *Scenario-Based Design: Envisioning Work and Technology in System Development*. New York: John Wiley & Sons, Inc.
- Carroll, J.M., Rosson, M.B., Chin, G. and Koeneman, J. (1997). Requirements development: Stages of opportunity for collaborative needs discovery. In *Proceedings of DIS 1997*. ACM Press, New York, pp. 55–64.
- Chamberlain, A., Crabtree, A., Rodden, T., Jones, M. and Rogers, Y. (2012). Research in the wild: Understanding ‘in the wild’ approaches to design and development. In *Proceedings of DIS 2012*. ACM Press, New York, pp. 795–796.
- Chapanis, A. (1959). *Research Techniques in Human Engineering*. Baltimore, MD: The John Hopkins Press.
- Clarkson, P.J., Coleman, R., Hosking, I. and Waller, S. (2007). *Inclusive Design Toolkit*. Cambridge Engineering Design Centre, University of Cambridge, Cambridge.
- Cockton, G. (2004). Value-centred HCI. In *Proceedings of NordiCHI 2004*. ACM Press, New York, pp. 149–160.
- Collins, M. (1986). Sampling. In R.M. Worcester and J. Downham (eds.), *Consumer Market Research Handbook*. McGraw-Hill Publishing, North-Holland.
- Comstock, E. (1983). Customer installability of computer systems. In *Proceedings of Human Factors Society, 27th Annual Meeting*. Santa Monica, CA: Human Factors Society.
- Crabtree, A., Rouncefield, M. and Tolmie, P. (2012). *Doing Design Ethnography*. Springer, Springer-Verlag, London.
- Crozier, R. (1994). *Manufactured Pleasures: Psychological Responses to Design*. Manchester University Press.
- Csikszentmihalyi, M. (1991). Design & order in everyday life. *Design Issues*. Manchester, 8, 26–34.
- Curtis, P., Heiserman, T., Jobusch, D., Notess, M. and Webb, J. (1999). Customer-focused design data in a large, multi-site organisation. In *Proceedings of CHI 1999*. ACM Press, New York, pp. 608–615.
- Desmet, P.M.A. and Hekkert, P.P.M. (2002). The basis of product emotions. In W.S. Green and P.W. Jordan (eds.), *Pleasure with Products: Beyond Usability*. London, U.K.: Taylor & Francis Group.
- Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics* 35, 1297–1326.
- Dittmar, H. (1992). *The Social Psychology of Material Possessions: To Have Is to Be*. Hemel Hempstead, U.K.: Harvester-Wheatsheaf.
- Druin, A. (1999). Cooperative inquiry: Developing new technologies for children with children. In *Proceedings of CHI 1999*. ACM Press, New York, pp. 592–599.
- Druin, A. (2002). The role of children in the design of new technology. *Behaviour & Information Technology* 21(1), 1–25.
- Dunne, A. and Raby, F. (2001). *Design Noir: The Secret Life of Electronic Objects*. Basel, Switzerland: Birkhauser.
- Durrant, A., Frohlich, D., Sellen, A. and Lyons, E. (2009). Home curation versus teenage photography: Photo displays in the family home. *International Journal of Human-Computer Studies* 67(12), 1005–1023.
- EGgen, B., Hollemans, G. and van de Sluis, R. (2003). Exploring and enhancing the home experience. *Cognition Technology and Work* 5, 44–54.
- Faulkner, X. (2000). *Usability Engineering*. Palgrave Macmillan, Basingstoke.
- Fulton, J. and Stroud, P.G. (1981). Ergonomic design of automatic ticket barriers for use by the travelling public. *Applied Ergonomics* 12(4), 203–207.
- Fulton Suri, J. (1993). User trials for information graphics: Replacing designers’ assumptions with feedback from users. *Information Design Journal* 7(2), 99–104.
- Fulton Suri, J. (2001). The ergonomics society – the society lectures 1999: The next 50 years: Future challenges and opportunities for empathy in our science. *Ergonomics* 44, 1278–1289.
- Fulton Suri, J. (2002). Whether to measure pleasure or just tune in. In W.S. Green and P.W. Jordan (eds.), *Pleasure with Products: Beyond Usability*. London, U.K.: Taylor & Francis Group.
- Fulton Suri, J. and Marsh, M. (1999). Scenario building as an ergonomics method in consumer product design. *Applied Ergonomics* 31(2), 151–157.
- Gaver, W., Boucher, A., Pennington, S. and Walker, B. (2004). Cultural probes and the value of uncertainty. *Interactions* 11(5), 53–56.
- Gaver, W. and Dunne, A. (1999). Projected realities: Conceptual design for cultural effect. In *Proceedings of CHI 1999*. ACM Press, New York, pp. 600–607.

- Gilmore, D.G. (2002). Understanding and overcoming resistance to ethnographic research. *Interactions* 9(3), 29–35.
- Goffman, E. (1971). *The Presentation of Self in Everyday Life*. Harmondsworth, U.K.: Pelican.
- Gould, J.D. (1988). Designing for usability: The next iteration is to reduce organizational barriers. In *Proceedings of 32nd Annual Meeting of the Human Factors Society*, Santa Monica, CA, pp. 1–9.
- Greenbaum, J. and Kyng, M. (eds.). (1991). *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, MI: Lawrence Erlbaum Associates.
- Greenberg, S. and Buxton, W. (2008). Usability evaluation considered harmful (some of the time). In *Proceedings of CHI 2008*. ACM Press, New York, pp. 111–120.
- Grudin, J. (1990). The computer reaches out: The historical continuity of interface design. In *Proceedings of CHI 1990*. ACM Press, New York, pp. 261–268.
- Hanington, B. and Martin, B. (2012). *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Beverly, MA: Rockport Publishers.
- Harboe, G., Minke, J., Ilea, I. and Huang, E. (2012). Computer support for collaborative data analysis: Augmenting paper affinity diagrams. In *Proceedings of CSCW 2012*. ACM Press, New York, pp. 1179–1182.
- Harper, R., Rodden, T., Rogers, Y. and Sellen, A. (2008). *Being Human: Human-Computer Interaction in the Year 2020*. Microsoft Research Ltd. Cambridge.
- Hasdogan, G. (1996). The role of user models in product design for the assessment of user needs. *Design Studies* 17, 19–33.
- Heath, C., Hindmarsh, J. and Luff, P. (2010). *Video in Qualitative Research*. London, U.K.: Sage Publications Ltd.
- Heath, C. and Luff, P. (1991). Collaborative activity and technological design: Task coordination in London underground control rooms. In *Proceedings of ECSCW 1991*. Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 65–80.
- Holtzblatt, K., Burns Wendell, J. and Wood, S. (2004). *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. San Francisco, CA: Morgan Kaufmann.
- Houde, S. and Hill, C. (1997). What do prototypes prototype? In M. Helander, T. Landauer, and P. Prabhu (eds.), *Handbook of Human-Computer Interaction*, 2nd edn. Amsterdam, the Netherlands: Elsevier Science B. V.
- Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B., Druin, A., Plaisant, C., Beaudouin-Lafond, M. et al. (2003). Technology probes: Inspiring design for and with families. In *Proceedings of CHI 2003*. ACM Press, New York, pp. 17–24.
- IDEO. (2003). *IDEO Method Cards*. Palo Alto, CA: IDEO.
- Isbister, K., Höök, K., Sharp, M. and Laaksolahti, J. (2006). The sensual evaluation instrument: Developing an affective evaluation tool. In *Proceedings of CHI 2006*. ACM Press, New York, pp. 1163–1172.
- ISO 9241 (Part 210). (2010). *Ergonomics of Human-System Interaction: Human-Centred Design for Interactive Systems*. Geneva: International Organisation for Standardisation.
- Joe, P. (1997). Scenarios as an essential tool: Stories for success. *Innovation Quarterly Journal of the Industrial Designers Society of America* Fall, 20–23.
- Jordan, B. and Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences* 4(1), 39–103.
- Jordan, P.W. (2000). *Designing Pleasurable Products*. London, U.K.: Taylor & Francis Group.
- Kantowitz, B.H. (1992). Selecting measures for human factors research. *Human Factors* 34, 387–398.
- Karat, C.M., Campbell, R. and Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Proceedings of CHI 1992*. ACM Press, New York, pp. 397–404.
- Kemp, J.A.M. and van Gelderen, T. (1996). Co-discovery exploration: An informal method for the iterative design of consumer products. In P.W. Jordan, B. Thomas and B.A. Weerdmeester (eds.), *Usability Evaluation in Industry*. London, U.K.: Taylor & Francis Group.
- Kensing, F. and Madsen, K. (1991). Generating visions: Future workshops and metaphorical design. In J. Greenbaum and M. Kyng's (eds.), *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, MI: Lawrence Erlbaum Associates.
- Kirk, D.S. and Sellen, A. (2010). On human remains: Values and practice in the home archiving of cherished objects. *ACM Transactions on Computer-Human Interaction* 17(3), Article 10, 43.
- Kirk, D.S. and Stanton Fraser, D. (2006). Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of CHI 2006*. ACM Press, New York, pp. 1191–1200.
- Kolko, J. (2011). *Thoughts on Interaction Design*. Burlington, MA: Morgan Kaufmann.
- Kuniavsky, M., Goodman, E. and Moed, A. (2012). *Observing the User Experience: A Practitioner's Guide to User Research*, 2nd edn. Waltham, MA: Morgan Kaufmann.



- Lang, A.R., Martin, J.L., Sharples, S. and Crowe, J.A. (2013). The effect of design on the usability and real world effectiveness of medical devices: A case study with adolescent users. *Applied Ergonomics* 44(5), 799–810.
- Langford, J. and McDonagh, D. (2003). *Focus Groups: Supporting Effective Product Development*. London, U.K.: Taylor & Francis Group.
- Lavender, S.A., Hedman, G.E., Mehta, J.P., Reichelt, P.A., Conrad, K.M. and Park, S. (2014). Evaluating the physical demands on firefighters using hand-carried stair descent devices to evacuate mobility-limited occupants from high-rise buildings. *Applied Ergonomics* 45(3), 389–397.
- Lazar, J., Feng, J.H. and Hochheiser, H. (2009). *Research Methods in Human-Computer Interaction*. Chichester: John Wiley & Sons.
- Lindley, S., Meek, S., Sellen, A. and Harper, R. (2012). “It’s simply integral to what I do”: Enquiries into how the web is weaved into everyday life. In *Proceedings of WWW 2012*. ACM Press, New York, pp. 1067–1076.
- Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.
- Mackay, W.E., Guindon, R., Mantel, M.M., Suchman, L. and Tatar, D.G. (1988). Video: Data for studying human-computer interaction. In *Proceedings of CHI 1988*. ACM Press, New York, pp. 133–137.
- Mackay, W.E., Ratzert, A.V. and Janecek, P. (2000). Video artefacts for design: Bridging the gap between abstraction and detail. In *Proceedings of DIS 2000*. ACM Press, New York, pp. 72–82.
- Mäkelä, A. and Fulton Suri, J. (2001). Supporting users’ creativity: Design to induce pleasurable experiences. In *Proceedings of International Conference on Affective Human Factors Design*. London, U.K.: Asean Academic Press.
- Mancini, C., Rogers, Y., Bandara, A., Coe, T., Jedrzejczyk, L., Joinson, A., Price, B., Thomas, K. and Nuseibeh, B. (2010). Contravision: Exploring users’ reactions to futuristic technology. In *Proceedings of CHI 2010*. ACM Press, New York, pp. 153–162.
- McCarthy, J. and Wright, P.C. (2004). *Technology as Experience*. Cambridge, MA: MIT Press.
- McClelland, I.L. (1984). Evaluation trials and the use of subjects. In E.D. Megaw (ed.), *Contemporary Ergonomics 1984: Proceedings of the Ergonomics Society Conference*. London, U.K.: Taylor & Francis Group.
- Meister, D. (1986). *Human Factors Testing and Evaluation*. Amsterdam, the Netherlands: North-Holland.
- Miller, H. (1997). The social psychology of objects. In *Understanding the Social World Conference*. Nottingham Trent University, Nottingham.
- Moggridge, B. (1993). Design by story-telling. *Applied Ergonomics* 24(1), 15–18.
- Moggridge, B. (2007). *Designing Interactions*. Cambridge, MA: MIT Press.
- Monk, A. and Howard, S. (1998). Methods & tools: The rich picture: A tool for reasoning about work context. *Interactions* 5(2), 21–30.
- Monk, A., Wright, P., Haber, J. and Davenport, L. (1993). *Improving Your Human Computer Interface: A Practical Technique*. New York: Prentice-Hall.
- Moore, P.A. (1985). *Disguised!* Waco, TX: World Books.
- Moran, T. (2002). Everyday adaptive design. In *Proceedings of DIS 2002*. ACM Press, New York, pp. 13–14.
- Muller, M.J., Halswanger, J.H. and Dayton, T. (1997). Participatory practices in the software lifecycle. In M. Helander, T.K. Landauer and P. Prabhu (eds.), *Handbook of Human-Computer Interaction*. Amsterdam, the Netherlands: Elsevier.
- Muller, M.J., Tudor, L.G., Wildman, D.M., White, E.A., Root, R.W., Dayton, T., Carr, R., Diekmann, B. and Dykstra-Erickson, E.A. (1995). Bifocal tools for scenarios and representations in participatory activities with users. In J. Carroll (ed.), *Scenario-Based Design for Human Computer Interaction*. New York: Wiley.
- Nielsen, J. (1994). Usability laboratories. *Behaviour and Information Technology* 13, 1–2.
- Norman, D.A. (2010). *Living with Complexity*. Cambridge, MA: The MIT Press.
- O’Brien, D.D. (ed.). (1982). *Design Methods: Seminar on Control Room Design*. Lancashire Constabulary HQ, Lancashire.
- Olmsted-Hawala, E., Murphy, E., Hawala, S. and Ashenfelter, K. (2010). Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In *Proceedings of CHI 2010*. ACM Press, New York, pp. 2381–2390.
- Oppenheim, A.N. (1999). *Questionnaire Design, Interviewing and Attitude Measurement*. London, U.K.: Pinter Pub.
- Palen, L. and Salzman, M. (2002). Voice-mail diary studies for naturalistic data capture under mobile conditions. In *Proceedings of CSCW 2002*. ACM Press, New York, pp. 87–95.
- Pedersen, J. and Buur, J. (2000). Games and movies: Towards innovative co-design with users. In S.A.R. Scrivner, L.J. Ball and A. Woodcock (eds.), *Collaborative Design*. London, U.K.: Springer.
- Peters, T. (Summer, 1998). Design is IT! *Design Management Institute Journal* 9(3), 14–19.

- Peters, T. (Winter, 2000). Design as advantage No. 1: The design + identity 50. *Design Management Institute Journal* 11(1), 10–17.
- Pettrelli, D., Whittaker, S. and Brockmeier, J. (2008). Autotopography: What can physical mementos tell us about digital memories? In *Proceedings of CHI 2008*. ACM Press, New York, pp. 53–62.
- Poggenpohl, S.H. (2002). Design moves: Approximating a desired future with users. In J. Frascara (ed.), *Design and the Social Sciences*. London, U.K.: Taylor & Francis Group.
- Poynter, R. (2010). *The Handbook of Online and Social Media Research: Tools and Techniques for Market Researchers*. Chichester: John Wiley & Sons.
- Randall, D., Harper, R. and Rouncefield, M. (2008). *Fieldwork for Design*. Springer Verlag GmbH.
- Rea, L.M. and Parker, R.A. (1997). *Designing and Conducting Survey Research: A Comprehensive Guide*. San Francisco, CA: Jossey-Bass.
- Rogers, Y., Sharp, H. and Preece, J. (2011). *Interaction Design: Beyond Human-Computer Interaction*, 3rd edn. Chichester: John Wiley & Sons.
- Rosson, M.B. and Carroll, J.M. (2002). *Usability Engineering: Scenario Based Development of Human-Computer Interaction*. San Francisco, CA: Morgan Kaufman Publishers.
- Rubinstein, R. and Hersh, H.M. (1984). *The Human Factor*. Digital Equipment Corporation/Digital Press, Maynard, MA.
- Sanders, E.B.N. (2002). From user-centered to participatory design approaches. In J. Frascara (ed.), *Design and the Social Sciences*. London, U.K.: Taylor & Francis Group.
- Sanders, E.B.N. and William, C. (2001). Harnessing people's creativity: Ideation and expression through visual communication. In J. Langford and D. McDonagh (eds.), *Focus Groups: Supporting Effective Product Development*. London, U.K.: Taylor & Francis Group.
- Sellen, A.J., Murphy, R. and Shaw, K.L. (2002). How knowledge workers use the web. In *Proceedings of CHI 2002*. ACM Press, New York, pp. 227–234.
- Serpiello, N. (2001). *Picture This: Collage as a Human Centered Research Method for Product*. HFES Consumer Product Technical Group.
- Shneiderman, B. (2003). *Leonardo's Laptop: Human Needs and the New Computing Technologies*. Cambridge, MA.
- Sommerville, I. (2010). *Software Engineering*, 9th edn. Pearson, Harlow.
- Spencer, R. (2000). The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company. In *Proceedings of CHI 2000*. ACM Press, New York, pp. 353–359.
- Stanton, N.A. and Young, M. (1998). Ergonomics methods in consumer product design and evaluation. In N. Stanton (ed.), *Human Factors in Product Design*. Taylor & Francis Group, pp. 21–52.
- Sterling, B. (2009). Design fiction. *Interactions* 16(3), 20–24.
- Suchman, L. (1987). *Plans and Situated Actions: The Problem of Human Computer Communication*. Cambridge, U.K.: Cambridge University Press.
- Sun, X., Golightly, D., Cranwell, J., Bedwell, B. and Sharples, S. (2013). Participant experiences of mobile device-based diary studies. *International Journal of Mobile Human Computer Interaction* 5(2), 22.
- Tolmie, P., Pycock, J., Diggins, T., MacLean, A. and Karsenty, A. (2002). Unremarkable computing. In *Proceedings of CHI 2002*. ACM Press, New York, pp. 399–406.
- Tufte, E.R. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.
- Vermeeren, A., Law, E., Roto, V., Obrist, M., Hoonhout, J. and Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: Current state and development needs. In *Proceedings of NordiCHI 2010*. ACM Press, New York, pp. 521–530.
- Verplank, B., Fulton, J., Black, A. and Moggridge, B. (1993). Observation and invention: The use of scenarios in interaction design. In *CHI Tutorial*. Amsterdam, the Netherlands: ACM Press.
- Vines, J., Blythe, M., Dunphy, P., Vlachokyriakos, V., Teece, I., Monk, A. and Olivier, P. (2012). Cheque mates: Participatory design of digital payments with eighty somethings. In *Proceedings of CHI 2012*. ACM Press, New York, pp. 1189–1198.
- Virzi, R.A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors* 34, 457–468.
- Vredenburg, K., Isensee, S. and Righi, C. (2001). *User-Centered Design: An Integrated Approach*. New York, NY: Prentice Hall.
- Wong, Y.Y. (1992). Rough and ready prototypes: Lessons from graphic design. In *Proceedings of CHI 1992 Posters and Short Talks*. ACM Press, New York, pp. 83–84.



---

# 11 Inclusive Design and Design for Special Populations

*Edward Elton and Colette Nicolle*

## CONTENTS

|                                                            |     |
|------------------------------------------------------------|-----|
| Introduction.....                                          | 300 |
| Why Is Inclusive Design Important?.....                    | 301 |
| Population Case.....                                       | 301 |
| User Case .....                                            | 302 |
| Business Case.....                                         | 302 |
| Ethical Case.....                                          | 303 |
| Legal Case.....                                            | 303 |
| Fundamental Principles of Inclusive Design .....           | 304 |
| Product Interaction Capabilities.....                      | 304 |
| Product Demand.....                                        | 305 |
| Capability–Demand Relationship .....                       | 305 |
| Inclusive Design Approach .....                            | 306 |
| Non-Stigmatising Products .....                            | 307 |
| Qualitative Methods in Inclusive Design .....              | 308 |
| General Guidance for Recruitment and Data Collection ..... | 308 |
| Identifying an Appropriate Sample .....                    | 309 |
| Venues and Timescales.....                                 | 310 |
| Ethical Procedures.....                                    | 310 |
| Documents.....                                             | 310 |
| Guidance for Specific Methods.....                         | 311 |
| Questionnaires .....                                       | 311 |
| Interviews and Group Discussions .....                     | 312 |
| Observation in the Laboratory and the Real World.....      | 312 |
| Summary of Qualitative Methods in Inclusive Design .....   | 313 |
| Quantitative Methods in Inclusive Design .....             | 313 |
| Capability Data.....                                       | 313 |
| Capability Measures.....                                   | 314 |
| Objective Performance Measures.....                        | 315 |
| Case Study.....                                            | 316 |
| Analytical Design Metrics.....                             | 318 |
| Summary of Quantitative Methods in Inclusive Design .....  | 320 |

|                                                                  |     |
|------------------------------------------------------------------|-----|
| Environmental Context of Use in Inclusive Design.....            | 320 |
| Environmental Contextual Factors.....                            | 320 |
| Impact on Product Interaction.....                               | 321 |
| Physical Environmental Factors.....                              | 322 |
| Social Environment.....                                          | 323 |
| Other Contextual Factors.....                                    | 323 |
| Identifying Contextual Issues in the Design Process.....         | 324 |
| Summary of Environmental Context of Use in Inclusive Design..... | 325 |
| Summary and Concluding Remarks.....                              | 326 |
| References.....                                                  | 326 |

## INTRODUCTION

There are many different approaches to design. The approach selected is often dependent on the type of value the product and/or service intends to deliver to the end-users. This chapter describes the inclusive design approach, which aims to deliver ‘mainstream products and/or services that are accessible to, and usable by, people with the widest range of abilities within the widest range of situations without the need for special adaptation or design’ (BS 7000-6 2005). Accessibility and usability are the key criteria of this approach. Accessibility refers to allowing users access to the features of products and/or services through their sensory, physical and cognitive capabilities. Accessible design therefore focuses on the ‘principles of extending standard design to people with some kind of performance limitation to maximise the number of potential customers who can readily use a product, building or service’ (BS 7000-6 2005). Usability refers to the extent to which a system, product or service ‘can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use’ (ISO 9241-210 2010). In simple terms, accessibility means that a diverse range of users are able to perceive, understand and physically interact with the product, and usability refers to whether users can perform tasks with the product effectively, efficiently and with satisfaction.

This approach to design has been given many different names across the globe: in 1985, Ron Mace used the term *universal design* in the United States (Ostroff 2001); in the United Kingdom, the term *inclusive design* is often used, while in other parts of Europe, the term *design for all* is used. Whilst different terminology may be used, the fundamental principles of this approach remain the same – considering diversity in mainstream design. The case to support this design approach was outlined in a research paper written by Roger Coleman (1994) entitled ‘The case for inclusive design – an overview’. In essence, the paper highlighted that the issue of disability (i.e. having reduced functional capability due to impairment) was not being addressed in mainstream design. Most mainstream designs were being developed for young, able-bodied healthy individuals, leading to people being disabled by designs that did not take account of the wide range of human capabilities that exist within the population. In other words, failure to account for the reduced functional capabilities of people with impairments in design can result in their being unable to complete certain tasks/activities, that is, they become excluded as they do not have a sufficient level of capability to access the product/service. Products and services were being developed for disabled groups; however, these tended to be devices for specialist functions or contexts, such as hospital aids and appliances (Cassim et al. 2007). There was a widespread desire among the disabled community to be included within the mainstream of consumer society (Clarkson and Coleman 2010). Also, the number of people with disabilities within the population was growing due to a significant increase in older adults (+65 years) in society. Coleman (1994) was instrumental in promoting this inclusive design approach, an approach that takes into account the varying capabilities of people across the population in order to develop mainstream products and services that do not disable or exclude people.

In essence, inclusive design can be categorised as a specific type of human-centred approach to design. The inclusive design approach specifically focuses on understanding the needs, capabilities and attitudes of people who have some form of impairment and then applying this knowledge to mainstream design. Thus, ergonomics/human factors (with its goal of understanding people and applying it to design) play a significant role in the inclusive design approach. A number of ergonomic methods and techniques have either been developed from scratch or adapted in order to aid the development of inclusive products and/or services. Virtual and wearable simulators have been developed which re-create some of the effects of ageing and disability to support empathic modelling, ranging from the Third Age Suit (Hitchcock et al. 2001) and Loughborough University SKInS (Sensory and Kinaesthetic Interactive Simulations) (Cook et al. 2012), to Cambridge Simulation Gloves and Glasses (Goodman-Deane et al. 2013, Clarkson et al. 2013 at [www.inclusivedesigntoolkit.com](http://www.inclusivedesigntoolkit.com)). Specialist software has been developed such as HADRIAN (Human Anthropometric Data Requirements Investigation and Analysis) (Summerskill et al. 2009, Marshall et al. 2010) and the Exclusion Calculator (Clarkson 2007, Goodman-Deane et al. 2011), and task analysis has been used to conduct demand assessments (e.g. Dong et al. 2007). Websites such as DesigningWithPeople.org at the Helen Hamlyn Centre for Design, Royal College of Art ([www.designingwithpeople.rca.ac.uk/](http://www.designingwithpeople.rca.ac.uk/)) and the inclusive design toolkit, which contains the Exclusion Calculator and Impairment Simulator Software (Clarkson et al. 2013 at [www.inclusivedesigntoolkit.com](http://www.inclusivedesigntoolkit.com)), have been developed with the aim of providing the design industry with the necessary resources to work with older users and implement inclusive design in practice.

The information presented within this chapter should be used as a starting point for inclusive design. Research is being carried out all over the world in this area; encompassing all of it within one chapter is not possible. The material within this chapter is a compilation of our knowledge, research and experience along with information from publications that were felt to provide the reader with a sound understanding of the key principles and methods which can be employed when following an inclusive design approach.

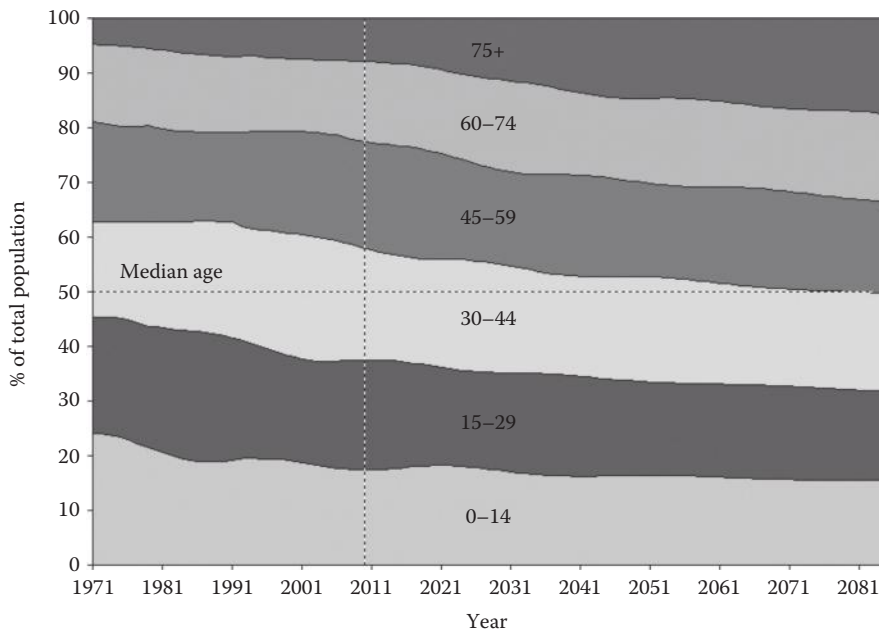
## WHY IS INCLUSIVE DESIGN IMPORTANT?

First, it is necessary to understand and establish the importance of inclusive design, for which a number of cases can be made. The strongest and most significant one remains the population case; however, there are also the business case, user case, ethical case and, increasingly, the legal case. Each of these cases will now be discussed.

### POPULATION CASE

Increased life expectancy and reduced birth rates have resulted in a greater proportion of older adults in society. For example in 1900, there were 3.1 million people aged over 65 in the United States; in 1990, there were 35 million; and in 2010, there were 40.3 million (U.S. Census Bureau 2011). According to Eurostat (2014), half of the European Union's population was older than 41.9 years in 2013 and those aged 65 years or over will account for 28.7% of the EU's population by 2080 (18.2% in 2013). Figure 11.1 shows how the age distribution of the United Kingdom is projected to change with this population increase and how the median age of the population is set to increase through this projected period.

Reductions in capabilities are typically experienced with age. These changes can be either due to the degenerative effects of ageing on the human body (i.e. changes to the form, structure and functioning of organisms within the human body) and/or the higher incidence of medical conditions with age, such as arthritis or cataracts. Thus, the number of people in society who are most likely to have reduced capabilities is significantly increasing, and population projections demonstrate that there is no sign of this slowing down in the future.



**FIGURE 11.1** Projected change of age distribution and median age in the United Kingdom. (Source: Office for National Statistics, 2011, re-used under the terms of the Open Government Licence.)

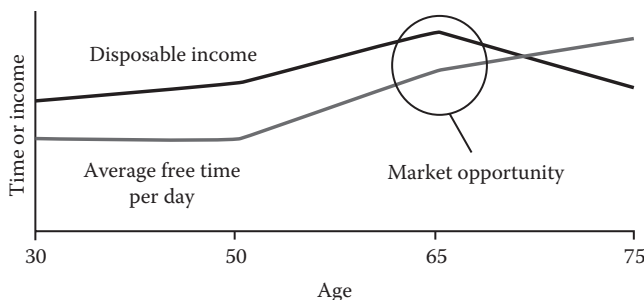
## USER CASE

As the baby boomer generation (people born after the World War II, between 1946 and 1964, approximately 450 million people worldwide) begins to reach older adulthood, it is suggested that we will experience an older generation that is conceptually different to any other (Hudson and Gonyea 2012). The first generation to grow up in a consumer society, they expect their individual wants and needs to be satisfied and are also the most economically powerful section of the UK society, as well as being considered smart consumers, aware of choices in the marketplace and very sensitive to being patronised (Huber and Skidmore 2003). Baby boomers have learned to expect great things from design and technology, in particular wireless information and communication technologies which have now become part of their everyday lives (Morris et al. 2009). Therefore, a more demanding, potentially more knowledgeable generation of older users is beginning to emerge, which inevitably will have significant impact on their needs and wants for products and services.

## BUSINESS CASE

When designing and developing a new product or service, one of the key questions asked is whether there is a viable market sector for it. A number of factors are often considered when carrying out such an activity, for example current and predicted value of the market, number of consumers and spending power. As the population case shows, there has been and continues to be a rapid increase in the number of older adults within society. There is also evidence to suggest an increase in disposable income in later life.

More recent figures from the ONS (2012) demonstrate that the average disposable income for retired households was £17,700 in 2010/2011, over two and a half times higher in real terms than in 1977. Therefore, even though Figure 11.2 refers to disposable income figures from the 1990s, it would appear that such trends still hold true in today's marketplace, with disposable income tending to peak around the time of retirement. Thus, the older adult market segment offers a real market opportunity that should not be overlooked.



**FIGURE 11.2** Disposable weekly household income, average daily free time by age. Courtesy of Henley Centre for Forecasting (now called The Futures Company), Leisure tracking survey, Family expenditure survey. Leisure futures, 1996, Vol. 3, pp. 7–9, London: U.K. (Adapted image accessed February 2014 from Clarkson et al. 2013 at [www.inclusivedesigntoolkit.com](http://www.inclusivedesigntoolkit.com).)

An older person's actual spending power will of course depend on so many other financial commitments or circumstances, for example whether they are now 'empty nesters' with more money to spend on products now that their children have moved away from home. Nevertheless, product developers should not overlook this spending power as a result of the 'grey pound' – designing and developing products and services for the older adult market is and should continue to be a viable option.

There is also another business case to consider for inclusive design. The inclusive design approach focuses on extending standard design to people with some kind of capability limitation; this process not only makes the product and/or service easier to use for the people with the impairment(s), but may also do so for others who possess a greater level of capability. Thus, through broadening the limits of a design, a greater number of people are able to use it. This allows businesses to take a single product and/or service and market it towards a wider market segment of consumers. However, it is important to understand that markets are not just segmented by age and capability; factors such as socio-economic status, geographic location, personality, motives and lifestyle are also used. Segmenting markets and positioning products are outside the scope of this chapter; however, the reader should be aware of other issues which impact upon design.

## ETHICAL CASE

Whilst there are many positive cases that support the inclusive design approach, there is also the societal consequence of not adopting it. Failure to develop products and/or services in line with the needs and capabilities of older and disabled people could result in their becoming excluded from use and therefore losing their independence. More specifically, this type of exclusion occurs when an individual's level of capability is less than that required to access and use the product and/or service (see 'Capability–Demand Relationship' section for further details). Products and services enable people to maintain their independence. In particular, products play a crucial role in activities of daily living, such as eating, bathing, dressing and toileting, and instrumental activities of daily living (IADL), such as shopping, food preparation, housekeeping, laundry and using a phone. Having the ability to carry out these activities enables a person to live independently. Products that are designed well are one of the key contributing factors that allow older adults to complete such tasks and thus maintain their independence. However, if products are designed badly, then it is likely that such tasks will either become problematic or no longer possible. Inevitably, this will result in people losing their ability to live independently and become dependent on others to carry out everyday activities.

## LEGAL CASE

Legally, disability cannot be used as a reason to treat people unfairly, as enforced by the Americans with Disabilities Act (U.S. Department of Justice 1990) and then by the United Kingdom's

Disability Discrimination Act, which later became the Equality Act (U.K. Government Equalities Office 2010) and now covers age and other characteristics – so the scope of legislation is moving on. We also now have, in addition to a number of EU directives, the United Nations Convention on the Rights of Persons with Disabilities (United Nations 2008), which has as its main objective to promote, protect and ensure the full and equal enjoyment of all human rights and freedoms by all. Accessibility is covered by Article 9 of this Convention, that is, that people with disabilities must have access to transportation, places and information, and not be refused access because of a disability. However, none of these acts and directives specifies that products have to be designed inclusively – they have to be safe and fit for purpose but do not have to be accessible. In fact, many people suggest that legislation, which would make inclusive design compulsory, is impractical and unlikely, and that a code of practice is more likely to be obeyed in spirit than in law (Rica 2001). Inclusive design is, however, on the agenda. For example, ANEC (the European consumer voice in standardisation at [www.anec.eu](http://www.anec.eu)) defends consumer interests in the process of standardisation and certification on aspects of products and services, including not just safety, performance and quality, but also accessibility. Its working groups, for example on design for all, have put inclusive design on several agendas, and the general direction of these initiatives looks promising. The European Commission is also expected to propose a European Accessibility Act in order to make sure that products and services, public buildings and spaces become more accessible to all citizens. This will be welcomed news as it will strengthen and extend existing initiatives, not just at European level but across the global economy. However, in addition to the ‘big stick’ approach, designers and developers need to be provided with ergonomic advice and methods in a suitable form that they can use to achieve more inclusive designs. This aspect of the inclusive design approach will be considered later in this chapter.

## FUNDAMENTAL PRINCIPLES OF INCLUSIVE DESIGN

This section describes the fundamental principles of inclusive design; this will help the reader to understand the key concepts which underpin this approach. The methods and techniques detailed later in this chapter focus on aiding people such as ergonomists, researchers, designers or engineers to embed these principles in design. The fundamental principles detailed within this section specifically focus on the design of products; however, these principles also hold true for many aspects of service and system design.

### PRODUCT INTERACTION CAPABILITIES

When interacting with a product, we typically use a combination of seven different human capabilities: these are vision and hearing (sensory capabilities), cognition (cognitive capabilities), dexterity, strength, locomotion, and reach and stretch (physical capabilities) (adapted from Clarkson et al. 2013). These capabilities are used to perceive, understand and physically interact with the product’s attributes in order to carry out the desired task. A description of each of these capabilities is provided in Table 11.1.

The capabilities we use and the frequency with which we use them to interact with products vary on a product-by-product basis. However, in most cases, vision and dexterity are the capabilities which are required most frequently when interacting with everyday products (Elton 2012). A user’s level of functioning (ability) along any of these capability dimensions can range from very high to extreme impairment (Johnson et al. 2009). Ageing generally causes a reduction in functioning along each of these capability dimensions to differing degrees, whereas a disability, such as being blind or deaf, can cause a significant reduction in functioning along just one capability dimension.



**TABLE 11.1**  
**Product Interaction Capabilities**

| Capability (Type)                   | Description                                                                                                                                                                                                                                       |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Vision</i> (sensory)             | The ability to sense colour and brightness of light to detect objects, discriminate between different surfaces or the detail on a surface.                                                                                                        |
| <i>Hearing</i> (sensory)            | The ability to interpret sound vibrations. People can identify simple sounds such as beeps and tones and complex sounds such as speech and music against differing background noise.                                                              |
| <i>Cognition</i> (cognitive)        | This refers to how a person understands and acts, that is, ability to process information, hold attention, store and retrieve memories, select appropriate responses and actions – language and social skills.                                    |
| <i>Locomotion</i> (physical)        | The ability to move around, bend down, climb steps and shift the body between standing, sitting and kneeling.                                                                                                                                     |
| <i>Reach and stretch</i> (physical) | The ability to put one or both arms out in front of the body, above the head or behind the back.                                                                                                                                                  |
| <i>Dexterity</i> (physical)         | The ability to use one's hands or the ability to manipulate objects with the hands (Trombly and Scott 1989).                                                                                                                                      |
| <i>Strength</i> (physical)          | The human ability to exert force against physical objects (Sanders and McCormick 1993). In relation to products, force is often exerted through our upper limbs and hands; thus, strength is closely interlinked with dexterity in this instance. |

## PRODUCT DEMAND

Products have a combination of attributes (buttons, handles, text, alerts, menus, etc.) which have to be perceived, understood and physically interacted with in order for the user to successfully interact with the product. These attributes place demands on user capabilities; in other words, they require a certain level of effort from the user in order for them to be interacted with successfully. For example, consider the numbers on a mobile phone keypad; users will require a certain level of visual capability in order for them to perceive the numbers successfully. Similarly, the cap on a bleach bottle will require a certain level of dexterity and strength to be opened, and a text alert will require a certain level of hearing ability in order for it to be heard. Demand can be thought of as the level of effort/capability a product requires of the user in order for it to be successfully used. It is important to remember that product demand is intrinsic to the product, and capabilities are intrinsic to the user – it is the relationship between these two factors which determines a user's product experience (this relationship is analogous to the relationship between demand and effort discussed in Chapter 18).

## CAPABILITY–DEMAND RELATIONSHIP

The relationship between the user's capabilities and the product demand determines the accessibility and usability (ease of use) of a product. There are three different consequences of the capability–demand relationship:

1. When our capability is *greater* than the product demand, the product is *easy to use*.
2. When our capability is *only just greater* than the product demand, then the product is *difficult* to use.
3. When our capability is *less* than the product demand, then we *cannot use* the product (EPSRC, Designing our Tomorrow project, ref: EP/H047670/1).

Completing a task with a product often requires multiple interactions that make a range of demands on our capabilities. For example, think about using your mobile phone to make a call: there are numerous steps involved in completing this task from retrieving the phone from your pocket/bag, entering the unlock code, navigating to the correct contact, dialling the number, holding a conversation, ending the call and finally placing the phone back in your pocket/bag. Each of these task steps places demands on a range of user capabilities, for example entering the unlock code places demands on a user's cognitive, dexterous and visual capabilities. The user's capabilities must be greater than the demand presented at each task step in order for the product to be accessible and usable. The following metaphor was created from a research project that the authors were involved with; the project (Designing Our Tomorrow, EPSRC ref: EP/H047948/1) aimed to engage key stage 3 school pupils (11–14 years) in the principles of inclusive design.

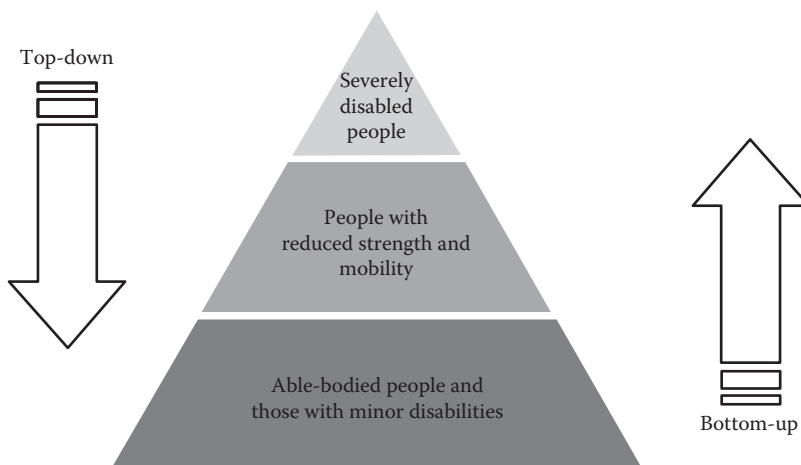
Think of the tasks steps you have to complete with a product as a series of hurdles. The height of the hurdle varies dependent upon the level of demand a task step presents. In order for the product to be easy to use, the user has to have the ability to clear each one of those hurdles. A product cannot be used when the user doesn't have the ability to clear a hurdle. (Ian Hosking, University of Cambridge, 2011)

As ergonomists, designers, engineers, etc., we can change the demand of a product (height of the hurdle) through varying the product characteristics. Products which are less demanding, that is require less ability/effort to use, are generally usable by more people – making life easier for everyone. Further information on the relationship between capability and demand can be found in Clarkson and Keates (2003) and Persad et al. (2006).

### INCLUSIVE DESIGN APPROACH

When designing for inclusivity, there are two main approaches that can achieve this: (1) top-down or (2) bottom-up. Both of these approaches are based around Benktzon's (1993) user pyramid as shown in Figure 11.3.

Top-down is designing for the least functionally capable users (those with the most severe impairments), then trying to make the resulting product mainstream friendly. Bottom-up is taking a mainstream product and pushing the boundaries of the design to include as many potential users as possible.



**FIGURE 11.3** The user pyramid – a map of disability across the population. (From Benktzon, M., *Appl. Ergon.*, 24(1), 19, 1993; Adapted by Keates, S. and Clarkson, P.J., *Countering Design Exclusion: An Introduction to Inclusive Design*, Springer-Verlag, London, U.K., 2004.)

One of the main problems of adopting a top-down approach is that the products may appeal/be sold only to a small target group and thus fail to penetrate the mainstream market (Keates and Clarkson 2004). Thus, adopting such an approach to design and develop inclusive products may prove impractical and not economically viable. However, a bottom-up approach offers greater potential for commercially successful products, as starting at the bottom of the pyramid and expanding upwards means large numbers of people are included. Also, users have a desire to comply with the norm – ‘I want what everyone else has, I don’t want to be different’ (Keates and Clarkson 2004). Thus, a bottom-up approach would appear to be the most practical and commercially viable way of achieving inclusivity in design.

The bottom-up approach is commonly known as the inclusive design approach. Following an inclusive design approach will help to ensure the product is accessible and usable by as many people as *reasonably* possible. Trying to make a product usable by absolutely everyone is a difficult, if not impossible, task. However, we can aim to exclude as few people as possible, and the objective should be that an inclusively designed product should exclude only those end-users whom the product requirements exclude (Keates et al. 2002). There will also always be people who will require an assistive device or another person to facilitate access, so we must not overlook the possible need for assistive products and technologies for users at the top of the user pyramid.

## NON-STIGMATISING PRODUCTS

Whilst inclusive products need to excel from both an accessibility and usability point of view, they also need to be desirable, too. Older and disabled users do not want to be seen to be different from the rest of society; they certainly do not want undesirable stigmatising products that are socially unacceptable (see ‘User Case’ section). An interesting observation once carried out by the authors saw an elderly participant, who suffers from Parkinson’s disease, remove the rubber grips from his ‘OXO Good Grips’ cutlery before eating. When asked why these had been removed, the participant replied ‘...I don’t need them...I’m not that disabled yet...’. Goddard and Nicolle (2012) investigated what was considered good design by older users. Results from this research showed that many older people would prefer, ideally, to have products that were not just easy to use but also looked good – but importantly, only if usability has been delivered. This idea links to Maslow’s (1987) hierarchy of needs, that is once the physiological and safety needs at the lower levels have been satisfied, emphasis can shift to needs at the higher levels, that is, towards more emotional attributes, for example desirability of a product. Jordan (2000) adapted Maslow’s model to a human factors perspective, creating a new three-level model with ‘functionality’ at the lowest level, ‘usability’ in the middle level and ‘pleasure’ at the highest level.

Addressing pleasure and desirability in design requires consideration of the less tangible human factors; it often involves tapping into people’s emotions, that is, desires, hopes, dreams, fears, identity, self-expression, etc. Methods that can be used to achieve this can be found detailed in ‘Designing Pleasurable Products’ (Jordan 2000). A paper written by Crossley (2003) provides a case study of how this has been achieved in a commercial design project, although younger users are the focus of this example. Even after this holistic understanding has been achieved, there is then the process of accurately translating this information into a set of design characteristics, which can sometimes prove challenging. User involvement at this stage is essential in ensuring emotions get accurately translated into desirable product characteristics. Qualitative research methods are particularly useful in obtaining information on people’s emotions and aiding the translation process. Such methods are also useful for identifying and exploring real-world problems and evaluating solutions. The following section discusses the use of qualitative research methods when working with older and/or disabled people.

## QUALITATIVE METHODS IN INCLUSIVE DESIGN

A variety of qualitative methods and techniques can be used in the design of products, technologies and services. Descriptions and usage of these techniques are given elsewhere in this book. However, because of the broad range of user characteristics and requirements that need to be taken into account when designing inclusively, certain qualitative techniques may require a bit more thought and consideration of the difficulties older and disabled users may have. This section suggests tips and protocols when involving older and/or disabled people in data collection. These tips have been gathered through the literature and experience over the years working with more vulnerable users towards better design solutions.

It is often difficult for designers to anticipate/foresee the problems older and disabled people may face with products as the characteristics of such user groups are very different to those of designers, who are often young-middle-aged healthy males. Increasingly, the call for inclusive design has challenged the boundaries of user-centred design to consider wide-ranging and diverse user characteristics and capabilities. User-centred design methodologies like USERfit (Poulson et al. 1996), originally developed to apply user-centred design to the assistive technology sector, can be used in following an inclusive design process. USERfit provides a structured approach with a set of summary tools covering context of use and analysis of the user, activity and product. The key is to encourage the designer or developer to make design issues explicit and to ask questions at each stage of the design process, so coming closer to a design that will be usable by more people.

A wide range of methods, tools and techniques can be used during different phases of the design cycle and for different purposes. For example, interviews and observation can be used during the analysis and specification phase, but also during evaluation. Using a variety of methods, including questionnaires, interviews, group discussions, observation and user trials, will provide a richer understanding of users' requirements and preferences, which may be diverse and changeable over time.

We can look at a scenario to see where some guidance may prove useful in recruiting an appropriate sample and during design and evaluation.

You are developing a questionnaire survey to assess the difficulties that people with disabilities have when looking for a job. The questionnaire will be sent to people with a wide range of disabilities.

Or

You are organising some user trials and questionnaires to assess the usability of a new coffee machine. You want to ensure that you include the needs of as many people as possible and recruit participants from a wide range of impairment groups.

How would you ensure that you accommodate the needs of

- People with visual impairments?
- People with hearing impairments?
- People with physical impairments?
- People with cognitive impairments?
- Frail elderly people?

## GENERAL GUIDANCE FOR RECRUITMENT AND DATA COLLECTION

Recruiting a representative sample for your study is difficult at the best of times, but it is even more challenging to obtain a representative sample of users with different types of impairments which may lead to problems, for example, in reading small text or being able to fill in a questionnaire. Sometimes it is difficult to obtain precise and comprehensive information because the user has difficulty in communication or because the user is too fatigued or in pain to complete the interview

or experiment. There is also an even greater onus on the researcher to ensure that ethical procedures are always followed, for example when inviting more vulnerable people to take part in an interview, especially when informed consent is not fully understood. Therefore, use of some techniques may require expert advice, even to say that certain techniques are not appropriate with some individuals. The next section provides tips on using questionnaires, interviews, group discussions and observation with different impairment groups. The aim is to encourage researchers and designers to successfully include their needs to the greatest extent possible and choose the methods and tools which will be most suitable for use with persons who are older or disabled, as well as following acceptable protocols and ethical considerations. Given the ageing population, the recommendation is to always include people over the age of 65 to the greatest extent possible, as we should see this as an exercise in 'designing for our future selves' (Benktzon 1993, Coleman and Pullinger 1993, Coleman 2001).

In order to ensure your data collection is suitable for capturing information from a diverse range of users, care is needed in choice of your user sample, venue, timescales, ethical procedures and production of documents.

### **Identifying an Appropriate Sample**

Ideally, no user groups should be excluded, but with resource constraints, it is always necessary to make certain difficult decisions and trade-offs. For example, when testing the usability of a product, it is suggested that for each characteristic or ability which needs to be considered, a minimum of 12 people should be recruited (Yelding and Cassim 2007). The best possible choice of user groups should be guided by the difficulties that might be experienced by people with different types of impairments when using the product or system under consideration. For example, a product which uses a visual display containing text is likely to cause difficulties for those with poor sight or language skills, and a product with small buttons set closely together could lead to error or frustration for people with visual impairments, as well as those with upper limb impairments including arthritis or tremor. When recruiting users, it makes sense to consider the person's functional abilities, that is, the ability or inability of a person to carry out a task or function, rather than the medical reason for their inability to perform a task. For example, in research investigating the needs of drivers with disabilities (Nicolle and Burnett 2001), separate focus groups were conducted with older drivers and drivers with impairments such as mobility or hearing. Whether a user had lost their mobility through illness or accident, the end result was the same in that they were not able to use the primary controls of the vehicle with their lower limbs. The needs of an older person with mobility problems may well be different from a younger wheelchair user, so this must also be taken into account and may require two separate groups, depending on the focus of the evaluation and what information is required from the users.

It is also necessary to consider the conflicting needs of different user groups and the adverse effects and consequences that a feature may have on one group, which might at the same time be benefiting another group. Therefore, not only should the design consider the features of the display, but it should also consider an alternative mode or format of output. 'Flexibility in use' is just one of the seven principles of universal design that can lead to inclusivity (North Carolina State University 1997). It is also important to consider users with multiple impairments, which are particularly true of the older population, who may, for example, have problems with both vision and hearing. Hence, as noted earlier, it is recommended that older people are included in the design and evaluation process.

It may be helpful to use what could be termed a 'cascade model' in recruiting participants for any research study (Richardson et al. 1993). A cascade model involves starting with high-level experts and spreading down to reach a wider number of users. This model helps to overcome some of the problems of identifying and contacting a diverse user population, over a wide geographical area across different nations or cultures, with heterogeneous characteristics, and who may need some encouragement and motivation to participate. Nicolle et al. (1993) found this model helpful to structure the data collection when identifying the requirements of drivers with disabilities. First, a focus

group was held with leading experts in the field (e.g. from driving assessment centres), and they then identified groups of drivers with disabilities who could participate in more detailed interviews. These people then either helped to arrange, or participated in, field observations. The use of this cascade data collection approach resulted in 56 experts being involved in the study and about 50 interviews with drivers with disabilities across seven countries.

### **Venues and Timescales**

If users are invited to attend an interview or user trial, the investigator should carefully consider the choice of venue, the distance that the person may have to travel and if they need assistance or extra information. We may wish to consider arranging their participation during existing meetings/venues, for example local clubs, day centres. Entry to and exit from the building, as well as the facilities within it (including parking, lifts with no stairs to climb and toilets) all need to be accessible. The room should also be free from background noise and reverberations, which can be reduced by the use of carpets, curtains, plants, etc. (Barrett and Kirk 2000). For people with visual impairments, the researcher should be ready to show or describe the layout of the room and other facilities, and describe any visual aids or illustrations that have been prepared.

It is important to ensure that the interview does not last too long, preferably no longer than 1–1½ h, with a break to avoid fatigue effects, and remember that participants may need more time to process information and give responses. The investigator also needs to allow time for social interactions, but it is useful to establish ground rules and endeavour to bring people back to the research topic as quickly as possible if you go off track. If more time is required, you may need to divide the study into two sessions, but always be prepared to do so at short notice at the request of the participant if they feel tired. Particular needs of user groups should be considered beforehand, and individuals always asked whether they have any special requirements. If you are conducting the study in the person's own home, you should be careful not to rearrange items or furniture, but if you do, take the time to put everything back in its rightful place.

### **Ethical Procedures**

Your organisation is likely to have ethical protocols in place when involving older and disabled people in user research, but whether there is a formal structure or not, certain guidelines are essential, including confidentiality and anonymity of information, and ability to leave the study at any time without prejudice (see Chapter 1). You may need to submit your proposed research for ethical review, which is required for any research with human participants, with further considerations if your participants are over 65 or are vulnerable in any way. Of course, ethical procedures are there for the benefit of both the participant and the investigator: having a second researcher there in the room (for note-taking) or within earshot (in case someone falls ill) is always useful, but may be particularly advisable or required when participants are older or disabled.

Once the participant has received information about the study, he or she gives informed consent. Although this is no different from other research, the format, style of the documents and how the user interacts with them may be different with older or disabled participants. For example, for someone with a visual impairment using a screen reader, participant information could either be sent by email or read over the phone beforehand. Similarly, the user might like you to read the informed consent with them or for them, as well as guiding their hand for signature. Needless to say, informed consent and all other documentation must be designed to be as accessible as possible, and some general guidelines are given later in the text.

### **Documents**

For all documents, no matter who your users are, the language should be kept simple and clear, but even more considerations are needed if older or disabled people are taking part. The best advice is to be flexible, keep the text in electronic form and ask people if they require the documents in a different format, whether it be large print, audio, Braille, or whether they would like the text emailed



electronically or read out to them. But remember that only a very small proportion of people with sight loss are able to read Braille, so converting all your documents to Braille is unlikely to be necessary, unless this is specifically requested.

Providing documents in electronic format can mean they are accessible by all sorts of people with and without sight loss, through their own computer facilities (e.g. ability to increase font size) and assistive technologies (e.g. screen readers and refreshable Braille displays). In any case, it is always recommended to pilot your data collection tools with a selection of older and disabled users to ensure their accessibility and usability.

In all your materials, consider the use of images to complement the text, as these will support people with dyslexia or low literacy, although for people with visual impairments, you will need to also provide a text description of the image in either audible or electronic form.

The RNIB (2006) has published (in book form and online) a practical guide and reference to help make your information accessible. This, and experience in the field, have contributed to the following selection of guidelines, which are good human factors to enhance legibility for everyone, and therefore essential for inclusive design:

- For clear print, use minimum font size of 12 pt with good contrast, but preferably 14 pt (RNIB 2006).
- Avoid using ornate or stylised typefaces, italics, underlining, right-justified lines, all capital or all bold letters, glossy or show-through paper, and setting text over images, as such features can make reading text more difficult for someone with visual difficulties (RNIB 2006).
- Avoid distinguishing between symbols by colour alone – different shapes and sizes can also help to distinguish between them. In fact, it is important always to use colour wisely, never using it as the only way to convey or code information (e.g., ISO 20282-1:2006).
- Use high colour contrast, that is, a light colour on a dark background or vice versa, and avoid using two light colours together, for example light blue and light green. The highest contrast can be achieved through using black lettering on a white background, or vice versa (RNIB 2006); this typically gives a 90% contrast level (Elton 2012).
- When using colour, it is important to remain mindful of people who suffer from colour blindness, that is people who have difficulty distinguishing between different colours within the colour spectrum. The most common form of colour blindness is red–green colour blindness; this is where people have difficulty in discriminating between red and green colours. However, there is also blue–yellow and total colour blindness. Colour blindness affects about 8% of men and 1% of women; using high colour contrast will enable a person to more easily recognise differences in colours, distinguished by their intensities. As a general guideline, always use light colours against dark colours, or dark colours against light, and you will not go too far wrong. Although pastel shades may look pretty, if there is lack of contrast with the background or with adjoining symbols, they may be difficult to distinguish and so should be avoided (RNIB 2006 and Royal National Institute of Blind People at [www.rnib.org.uk](http://www.rnib.org.uk)).

## **GUIDANCE FOR SPECIFIC METHODS**

This section provides a few tips when using specific data collection methods and tools with older people or people with disabilities. The purpose is to ensure that users are not excluded from the data collection process due to a lack of the investigator's forethought and planning.

### **Questionnaires**

The use of postal questionnaires may overcome users' mobility problems, but they may lead to difficulties in access or understanding for people with visual, dexterity or cognitive impairments.

Consider all the recommendations for preparation and layout of documents given in this chapter, taking care with spacing, chunking information into small segments, headings, numbering and cues (e.g. bold or colour for emphasis). Although a font size of 14 pt is recommended (RNIB 2006), sometimes trade-offs may be necessary if you are photocopying large numbers of questionnaires, but you should follow this advice wherever possible since your users may have sight loss. For postal or self-completion questionnaires, some older or disabled people may need help in completing it, either in person or over the phone, or they may require it in a different format. If, however, someone helps them to complete it or answers on their behalf, care is needed in data interpretation since the answers may not be wholly the recipient's.

Even though the data will be anonymised, it is often useful to invite respondents to provide contact details if they would like to be involved in further research, emphasising that these would be kept separate and could not be linked to their individual data. In a major questionnaire study undertaken by the author, the majority of participants provided their name and either telephone number or email address, indicating that we could get back to them with either queries on their data or with further research questions. This also enables the researcher to build up a database of participants for future studies.

### **Interviews and Group Discussions**

When interviewing older or disabled people, whether individually or as a group, carefully consider the physical environment in which the interview will take place and follow all the advice on venues provided earlier in this chapter. If participants have hearing or speech/language impairments, speak clearly and slowly, facing the participants. If lip reading is needed, avoid bright light behind the speaker and ensure that the researcher and participants are facing each other. You may need to enlist the support of a sign language interpreter, trained interviewer or facilitator to interpret the responses. Speak clearly but not abnormally, use normal volume and provide visual aids where possible. Individual interviews are well suited for people with cognitive impairments, for example, to identify their everyday difficulties or where technology might be able to help them. Questions should be specific and simple, use examples and be prepared to explain things in different ways. It is also often appreciated by participants if you suggest that they may bring a friend along to make them feel more secure/at ease. In a group setting, the use of name cards is helpful to call on participants, and it is also useful to establish a cue to indicate when a person wishes to speak, for example raising the hand or waiting for name to be called, the type of cue depending, of course, on whether or not the participants have visual or hearing impairments.

Motivating participants to take part is always an issue; however, the opportunity to influence better design and provision of information to older and disabled people, combined with refreshments, is often sufficient incentive to attend (Barrett and Kirk 2000). Studies by Barrett and Kirk (2000) also recommend over-recruiting by about 20% to allow for people not showing up on the day and suggest that this is especially wise when planning focus groups and interviews with older and disabled people and their carers.

### **Observation in the Laboratory and the Real World**

Observation in both real-world and laboratory settings must consider all the issues given in this section; otherwise, it could either cut short the testing or eliminate certain users entirely. What people say they do (e.g. in an interview) might not be what they actually do in practice. Therefore, observing older or disabled people will not only reveal their difficulties, but also identify any coping strategies they may use in order to complete a task. For example, when interviewing an older person about making a hot drink, they may say they have no problem; however, observations may reveal the use of strategies such as filling the kettle with a small jug, using only a small light-weight kettle or even using a microwave instead of a kettle (e.g. Sims et al. 2012). Gathering information and understanding through both interview and observation will lead to richer data, and asking the user to also think aloud (verbal protocol) whilst interacting with a product will help the investigator learn even more

about the users and the tasks they are performing. These data can be analysed through task analysis techniques in order to understand the users' requirements and abilities, to identify the most difficult aspects of a task, what functions and design aspects they like and what operations they dislike or find most difficult. All these aspects are important to attain a more inclusive design, whether designing a new product, service or technology, or a new version of an existing one (Dong et al. 2007).

### **SUMMARY OF QUALITATIVE METHODS IN INCLUSIVE DESIGN**

The use of qualitative methods in inclusive design will help you understand the users' requirements, for example how they feel about current products and technologies, what they find easy or difficult to do and ideas for new design concepts. It is important to remember that what people say they do (e.g. in an interview) might not be what they actually do in practice, and using more than one method to answer these questions will lead to richer data. The basic guidance provided here is good human factors for everyone, but has been supplemented with extra consideration when involving older and disabled users. Qualitative data can help to answer these questions, but quantitative data on users' capabilities must also play a role in designing more inclusive products.

### **QUANTITATIVE METHODS IN INCLUSIVE DESIGN**

If we are to design inclusive products, we need to ensure that our designs take into account the reduced functional capabilities of older and disabled people. In order to achieve this, we require knowledge of their product interaction capabilities, that is, their visual ability, hearing ability, dexterous ability or strength. These data can be obtained from anthropometric datasets, capability surveys and published studies, and/or through directly measuring the capabilities of the intended user population. Over the past decade, capability data have also been used to calculate the inclusive merit of a product's design (i.e., the number of people included/excluded from using a particular product). Analytical design metrics which quantitatively assess the number of people able to use a product was found to be one of the key factors that would greatly assist the implementation of inclusive design in business (Gyi et al. 2000, Dong et al. 2003). In essence, this involves applying the capability–demand theory (see 'Fundamental Principles of Inclusive Design'), that is, determining the demand of a product and comparing it to a representative capability dataset, to determine who will and who will not have the capability to access the design. For example, a keypad for an ATM may require a user to be able to distinguish numbers that are 2.5 mm tall. A representative capability dataset can then be referred to in order to determine what percentage of the target population have this level of visual capability and are thus included. Therefore, the quantitative methods that are typically used for the purposes of inclusive design are (1) capability data, (2) capability measures and (3) analytical design metrics which calculate the inclusive merit of a design. The following section will provide an overview of each of these quantitative methods.

#### **CAPABILITY DATA**

In order to design inclusive products, knowledge is required of the full range of product interaction capabilities within the target population. This can be best captured through either of the following two sources of data:

1. Anthropometric datasets, which typically detail static (e.g. stature, weight, breadths, depths, circumferences) and dynamic measurements (e.g. range of motion of joints, strength, walking and balance performance measures) of the human body; the measurements are typically concerned with the physical parameters of the human body.
2. Capability surveys, which typically detail data on the functional capabilities of users (i.e. what people are capable of doing); these include measures of sensory, cognitive and physical capabilities.

Both data sources typically detail the body size and/or capability ranges of a representative sample of the population along a given dimension(s). These data are intended to contribute towards the design process to help ensure that the product being developed is accessible and usable for the intended users (Pheasant and Haslegrave, 2006). Over the past 35 years, several older and/or disabled adult anthropometric and capability datasets have been developed; the most useful for inclusive design include the following:

1. Older adult data (Smith et al. 2000)
2. Bodyspace (Pheasant and Haslegrave, 2006)
3. Humanscale (Diffrient et al. 1978, 1981a,b)
4. Great Britain Disability Follow-Up Survey (DFS) (Grundy et al. 1999)
5. Better Design Survey (Clarkson et al. 2010)

The DFS was initially undertaken to help plan welfare support; however, the data gathered in this survey has since been reported as being suitable for use as a design-relevant capability database, as it contains data on the widest range of capability domains at a level of specificity suitable for adaption to capability estimates for product interactions (Johnson et al. 2009). However, strictly speaking, this is not a quantitative method, as the data were gathered using self-report measures, that is, when an individual subjectively rates their own ability against a number of questions or rating scales; thus, the data are not numerical data. The most recent and relevant capability survey to be conducted is the 'Better Design Survey'. This survey was conducted in 2010 as part of the EPSRC funded i~design3 research project, to specifically gather population-based data on user abilities that can better inform product design. The survey can be accessed online from the UK Data Service website at [ukdataservice.ac.uk/](http://ukdataservice.ac.uk/) (enter the serial number '6997' in the search box on the homepage to locate the database). The survey details information and data on general health, vision, hearing, mobility, reach, dexterity, cognitive function, anthropometrics, demographics, technology experience and psychological resources. Although the Better Design Survey was intended as a pilot study, the results from its 362 participants remain the most design-relevant capability dataset to date.

Finally, capability data can be obtained from published research studies in the form of journal papers and online formal reports. It is unlikely that any one single paper will contain a breadth of data-detailing capability ranges across a number of capabilities. However, there are studies which will provide normative data on a specific capability dimension for older or disabled adults' abilities. For example, Desrosiers et al. (1995) produce normative reference data on the fine finger capability of 360 healthy adults aged 60 and over; Hirvelä et al. (1995) measured the corrected visual acuity at a distance of 6 m of 476 adults aged +70 years in Finland. Often, a thorough literature search of publication databases such as Science Direct, Ergonomics Abstracts or Google Scholar is needed in order to find such papers.

## CAPABILITY MEASURES

It may not always be possible to obtain the specific capability data you require from published sources of capability data. An alternative option is to therefore measure the desired capabilities of the population of interest. Gathering capability data can be a timely and costly exercise; however, it is highly likely that the time and effort involved in gathering the necessary capability data will be worthwhile in terms of the benefits your product will deliver to the end-users.

When measuring capability, there are a number of important factors to consider. This includes the types of measures you use, the granularity of the measurement, the reliability and validity of the chosen measure(s), how relevant the measure is (i.e. does it measure the same type of capability used when interacting with the intended design), the size of the measuring scale (i.e. is it too fine, just right or too gross) and the size and composition of the sample of users you intend to measure.

An example of how some of these factors were addressed in a recent study by the authors will be detailed in ‘Case Study’ in this section.

Capability data can be gathered using a number of different measures, which include the following:

- *Self-report measures*: When an individual subjectively rates their own ability against a number of questions or rating scales.
- *Proxy measures*: When another (authorised) person is asked to provide capability information for an individual who is not able to do so.
- *Objective performance measures*: When an individual’s ability is measured whilst carrying out a task. Tasks can be measured by metrics such as speed, accuracy or strength.

There are both advantages and disadvantages to using each of these measures in the measurement of capability to facilitate inclusive design. An extensive review of each of these measures can be found in Johnson et al. (2009); however, a brief summary of the conclusions made by Johnson et al. (2009) and other authors are provided:

1. Self-report measures can potentially provide access to physiological, environmental, cultural and attitudinal components of capability. However, responses can be significantly affected by a number of factors such as a person’s affective state, their educational, cultural, language and social differences and the participant’s own judgements.
2. Proxy report measures are beneficial when individuals have insufficient capability to complete a survey. This method has been shown not to be as accurate as self-report among older people.
3. Objective performance measures have been found to be good at assessing functional ability or capacity to perform a particular task. They have been argued as being more reliable, more sensitive to change and more able to accurately measure ability at higher levels. Performance measures are also more likely to reflect physiological aspects of capability and also have the advantage of providing direct measures of human response. See Chapter 24 of this Volume. (Sainio et al. 2006).

The following section will focus on objective performance measures as they specifically produce quantitative data.

### **Objective Performance Measures**

The granularity of the measurement is an important consideration when selecting an appropriate objective capability measure(s). Johnson et al. (2009) identify three levels of measures:

1. Component function – for example, measuring vision using an eye test or dexterity with a pegboard.
2. Broad activities – for example, replying to an SMS message, vacuuming a room, using an iron.
3. Specific tasks – for example, reading the LCD display on a phone, fastening a button, opening a jar.

The main distinction between activities and tasks is that activities measure people’s overall ability to interact with a product, as activities often require the use of multiple capabilities. For example, replying to a text message requires cognition, vision and dexterity. However, for tasks, the focus is on completion of a specific action, which often requires the use of just one type of capability for it to be completed. For example, reading an LCD on a phone predominantly requires vision.

When considering the granularity of measurement, it is important to consider what you are trying to achieve. If you wish to gather capability data that are accurate, reliable and valid, then objective performance measures are often the most suitable. If you are interested in gathering capability data that are directly relevant to the design of a particular product, then task measures are often most suitable. Johnson et al. (2009) argue that task performance measures provide detailed data on product interaction ability and thus are more directly relevant to designers compared to the data gathered from activity measures. Tenneti et al. (2012) also state that task performance measures provide the highest level of granularity in relation to capability measures with products. One of the issues with task and activity measures is the validity and generalisability of the data they gather. From a validity perspective, both tasks and activities may not provide a true measure of an individual's level of capability as participants may be able to make informed guesses as to how to complete a specific task based on their familiarity with the product's features or the type of layout being used. For example, participants may be set the task of reading a keypad on a mobile phone; familiarity with the keypad layout may mean they know the top left button has the number 1 on it without actually being able to see it. There are also issues regarding the generalisability of such capability data to different products, as the data gathered are specific to the product used. For example, if your participants are able to read numbers 2.5 mm tall at 90% contrast on an ATM keypad, it does not mean they can do the same on a mobile phone keypad as a number of variables will change, such as the purpose of the task, layout of the keypad, colours and materials used. In studies carried out by the authors (Elton 2012, Elton and Nicolle 2013, Elton et al. 2013), a combination of objective performance and task measures were used in order to obtain a reliable, valid and relevant measure of capability in relation to product interaction.

### Case Study

The authors conducted a study to determine older adults' visual acuity capabilities (i.e. the eye's ability to discriminate detail) under different everyday ambient illumination levels. The aim of the study was to generate context-specific capability data that could be applied to the design of everyday products so that their visual characteristics could be successfully perceived by older adults in everyday environmental conditions. Initially, a review of visual acuity component function measures was undertaken; logarithm of minimum angle of resolution (LogMAR) charts were identified from the literature as the visual acuity measure that would provide the most reliable and valid measure of this type of capability. A review of commercially available LogMAR acuity charts revealed that current LogMAR charts had only been designed to test distance visual acuity (e.g. 6 m) and not near visual acuity (i.e. short-sighted capability), which is what products typically place demands upon. Thus, near distance visual acuity charts were developed specifically for this study in order to obtain a measure of capability that was relevant to product interaction; they were developed to measure visual acuity at a distance of 1 m. The next factor to be considered was the contrast of the letters used on the charts: typically, black letters on a white background are used on visual acuity charts. Black lettering on a white background provides approximately 90% contrast; however, the contrast level used on everyday products varies significantly. It was therefore decided to measure visual acuity across a range of contrast levels; these were 90%, 70%, 50% and 30% contrast. Again, this was another factor which had to be considered in order to ensure the data gathered were of relevance to the design of products. The LogMAR acuity charts that were developed for this experiment are illustrated in Figures 11.4 through 11.7.

A range of product tasks which predominantly require visual acuity to complete were also used to measure visual acuity in this experiment. The tasks involved participants identifying symbols, words and/or sentences on each product; a variety of product characteristics were selected which ranged in both size and contrast. Also, where possible, unfamiliar visual characteristics were selected; this was to reduce the chance of participants being able to make an informed guess about what they were viewing. The purpose for selecting product tasks was to establish the generalisability (external validity) of the data obtained from the LogMAR acuity charts to the design of products.





FIGURE 11.4 Ninety per cent contrast LogMAR chart.



FIGURE 11.5 Seventy per cent contrast LogMAR chart.



FIGURE 11.6 Fifty per cent contrast LogMAR chart.



FIGURE 11.7 Thirty per cent contrast LogMAR chart.

In other words, if a person achieves a high score on the LogMAR charts, do they also perform well on the product tasks? If not, the data obtained from the LogMAR acuity charts could not reliably be applied to the visual characteristics of products.

Following the experiment, a correlation analysis was conducted on the LogMAR and product task data. The findings from the Spearman's rho ( $r_s$ ) correlations showed that, in each of the ambient illumination conditions tested (street lighting, in-house and overcast), a strong significant relationship ( $r_s = -0.6$ ) existed between ability on the LogMAR acuity charts and visual ability to complete product tasks. These results showed that, as a component function measure, LogMAR acuity charts are good predictors of older adults' visual ability with products. Thus, confidence can be had in the generalisability of the data produced from LogMAR acuity charts to the visual characteristics of products. For further detail about this study and the results, see Elton (2012) and Elton et al. (2013).

### ANALYTICAL DESIGN METRICS

Finally, there are analytical design metrics that can be used to quantitatively assess the inclusive merit of a product's design (i.e. the number of people included/excluded from using a particular product). Gyi et al. (2000) and Dong et al. (2003) identified such tools/metrics as greatly assisting the implementation of inclusive design in business. The analytical design metrics discussed in this section are online design tools that were developed based on the findings from these research studies. Prior to introducing the various tools, it is important to point out that the most valid technique for determining the inclusive merit of a design will always be to have users interacting with the product in question; however, a variety of constraints (i.e. time, cost, resources) mean this will not always be possible (Johnson et al. 2009), at least not to the extent that we would recommend. The following tools were therefore developed as a way of supporting designers to develop inclusive products.

The following three capability tools are currently available. These can be taken as good examples of tools that provide analytical inclusive design metrics:

1. HADRIAN (Human Anthropometric Data Requirements Investigation and Analysis) (see Figure 11.8)
2. Exclusion Calculator (part of the online Inclusive Design Toolkit, Clarkson et al. 2013 at [www.inclusivedesign toolkit.com](http://www.inclusivedesign toolkit.com)) (see Figure 11.9)
3. Context Calculator (see Figure 11.10)

HADRIAN is a software-based computer-aided design tool that builds upon the digital human modelling system SAMMIE (Summerskill et al. 2009, Marshall et al. 2010, 2015). HADRIAN can



**FIGURE 11.8** HADRIAN inclusive design tool (Marshall, Summerskill et al. 2010).

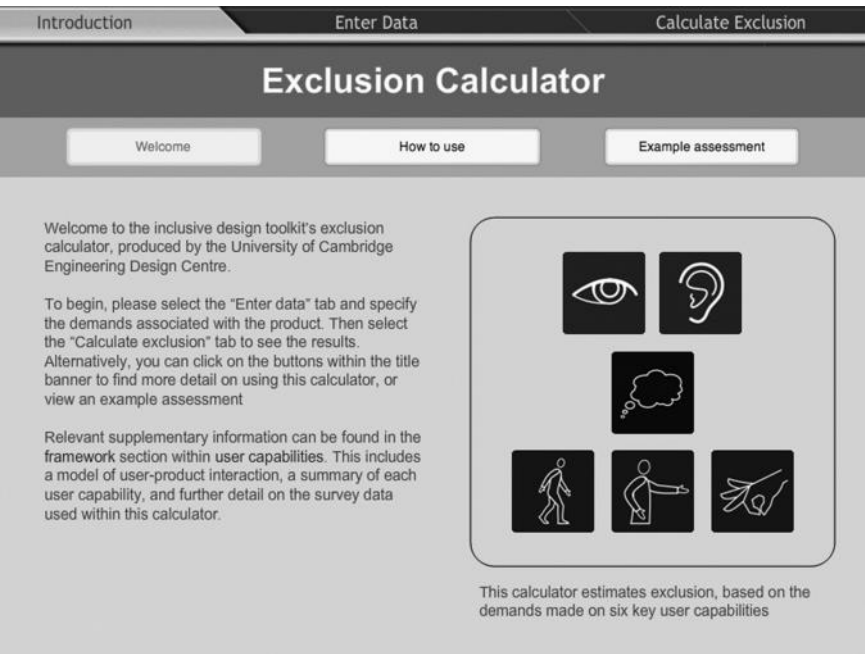


FIGURE 11.9 Exclusion Calculator at [www.inclusivedesigntoolkit.com](http://www.inclusivedesigntoolkit.com).

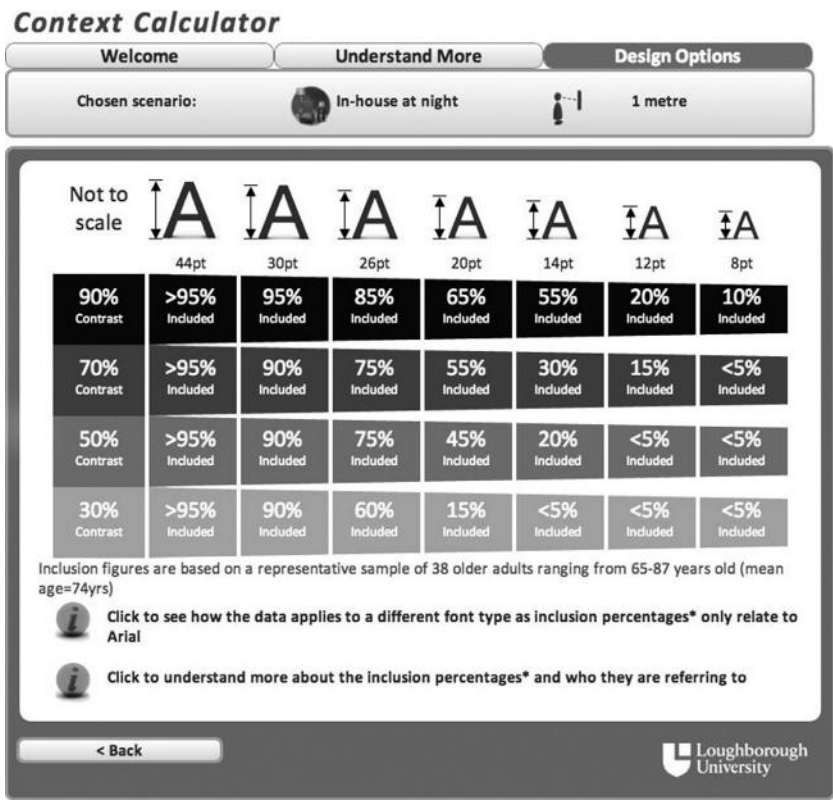


FIGURE 11.10 Context calculator.

be used to automatically assess/evaluate the inclusive merit of a proposed design (prepared as a CAD model). It compares the capabilities of 102 older and disabled adults to the specified tasks that are performed with the CAD model. The tool then highlights which users are unable to complete certain task stages based upon their ability. Features that cause difficulty and/or exclusion can then be modified to increase the inclusive merit of the product prior to manufacture.

The Exclusion Calculator is a web-based tool that is part of the online Inclusive Design Toolkit (Clarkson et al. 2013 at [www.inclusivedesigntoolkit.com](http://www.inclusivedesigntoolkit.com)). The Inclusive Design Toolkit is an online resource that aims to provide the design industry with the necessary resources to implement inclusive design. The Exclusion Calculator is designed to estimate the number of people within the population who would be excluded from using a particular product based on assessing the demands on each individual capability domain (Clarkson et al. 2013). Attributes that require the greatest level of capability (i.e., exclude the most people) can then be identified and refined so they are less demanding, thus decreasing the number of potential users being excluded.

The Context Calculator is the final web-based tool detailed in this section; it can be accessed from [www.lboro.ac.uk/departments/lds/research/groups/user-centred-design/](http://www.lboro.ac.uk/departments/lds/research/groups/user-centred-design/). This tool was developed by the authors (Elton and Nicolle) and is the first inclusive design tool which focuses on calculating exclusion caused by environmental context of use. It specifically focuses on the effects of everyday ambient illumination conditions on older adults' visual capabilities and the effects of an everyday cold temperature on older adults' dexterous ability. The tool details inclusion data on how readable different letter size/contrast combinations are in a range of everyday lighting conditions, as well as providing guidance on how to minimise the effects of an everyday cold temperature on product interaction.

## SUMMARY OF QUANTITATIVE METHODS IN INCLUSIVE DESIGN

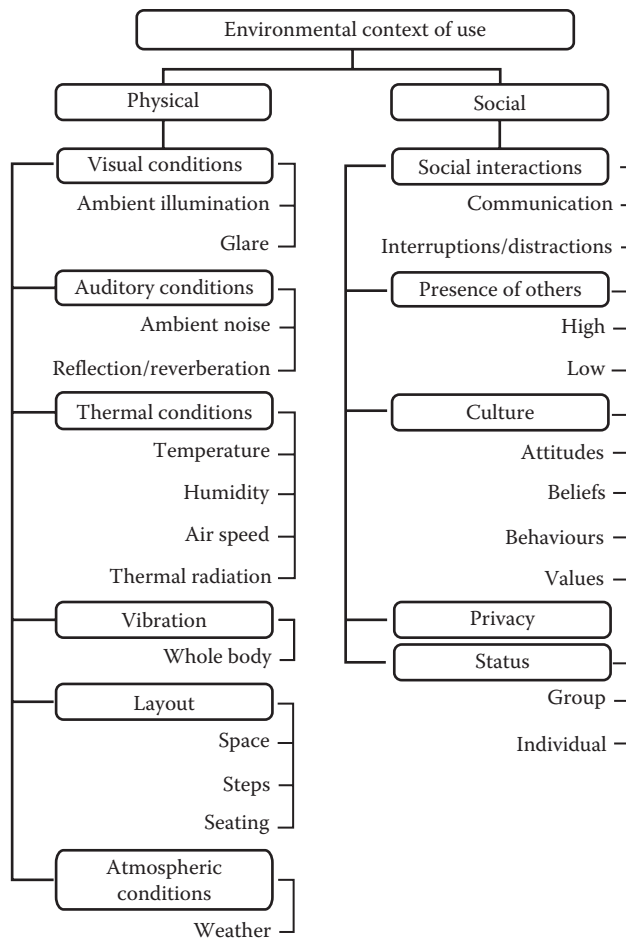
The use of quantitative methods in inclusive design will help you to design and develop products that take account of the reduced functional capabilities of older and disabled people. Data tools will help you to develop products which are less demanding and thus ultimately include more users. As well as considering capability data, it is also important to consider environmental context of use. This is particularly important in the area of inclusive design as everyday environmental conditions can significantly affect older adults' capabilities and thus affect their ability to interact with products. The final section of this chapter will discuss the importance of environmental context of use in inclusive design.

## ENVIRONMENTAL CONTEXT OF USE IN INCLUSIVE DESIGN

Environmental context of use refers to the physical and social environments in which a product is used (ISO 9241-11 1998, ISO 20282-1 2006). The 'physical environment' refers to the external surroundings and conditions, and the 'social environment' refers to the culture and the interaction of people that surround the user(s) (ISO 9241-11 1998, ISO 20282-1 2006). Put more simply, environmental context of use refers to the environmental circumstances in which a product interaction(s) takes place.

## ENVIRONMENTAL CONTEXTUAL FACTORS

Prior to discussing the impact of environmental context of use on product interaction, it is important to understand what environmental factors are deemed relevant to design. Following a comprehensive review of environmental context of use, Elton et al. (2008) and Elton (2012) identified a range of contextual factors associated with either the physical or social environment that were deemed relevant to design – these are represented in a conceptual framework of environmental context of use, as detailed in Figure 11.11.



**FIGURE 11.11** Environmental context of use framework. (From Elton, E., *Generating and translating context capability data to support the implementation of inclusive design within industry*, Doctoral thesis, Loughborough University, Loughborough, U.K., 2012.)

### IMPACT ON PRODUCT INTERACTION

Consideration of the environmental factors identified in Figure 11.11 is important when developing any product. It is particularly relevant when designing and developing inclusive products since environmental conditions can significantly reduce an older adult's product interaction capabilities, capabilities that are already impaired due to the effects of ageing and/or impairments. Another important factor to consider is that older adults are far more likely to be affected by environmental conditions due to their reduced functional capabilities (Iwarsson et al. 2007). This is due to the fact that anatomical and physiological changes that occur to the human body, as a result of ageing, make adapting and functioning under differing environmental conditions more problematic. For instance, there is a dramatic slowing in the eye's ability to adapt to the dark with ageing, which means objects are difficult to recognise at low illumination levels (Jackson et al. 1999). Also, ageing causes a degenerative loss in skeletal muscle mass (sarcopenia) and reductions in muscle fibre length resulting in reduced grip strength, mobility and time to exhaustion (Carmeli et al. 2003). Thus, the ability to overcome the effects of certain environmental conditions is significantly reduced. The overarching principle that applies here is that failure to consider the environmental context when designing inclusive products could easily result in a product causing difficulties to, or excluding, those intended to be included.

All of the identified environmental contextual factors have the potential to decrease a user's product interaction capabilities (e.g. vision, hearing, cognition, dexterity, reach and stretch, locomotion and strength) to some degree. A reduction in capability may lead to users becoming excluded as the demand of using a product, in a specific environment, may exceed the user's capability levels. The extent to which these environmental factors affect a person's capability is dependent upon such variables as exposure time, intensity/level and capability level. A selection of examples are now detailed to provide the reader with an understanding as to how these factors can impact on a user's capability and thus affect product interaction.

### Physical Environmental Factors

*Ambient illumination:* the human eye automatically adjusts itself to changes in light intensity.

At low levels of illumination, the eye loses its ability to perceive detail (visual acuity) and increases its ability to detect light. Elton et al. (2013) found that on average, older adults' (+65 years) visual acuity decreased by 1.6 times when under street lighting conditions (7.5 lx) compared to overcast conditions (6000 lx).

*Glare and reflection:* a situation where there are light levels, within the line of sight, that are relatively high compared with ambient light levels (see Chapter 25 of this volume). Glare can be experienced through a bright light being reflected from a product's surface into the field of view (Figure 11.12). Glare causes annoyance, discomfort and can significantly reduce a user's visual ability.

*Cold temperature:* can result in physiological changes to the joints and muscles within the hand and arm, which cause reductions in dexterous ability (Havenith et al. 1995). In particular, it can lead to reductions in fine finger dexterity; Elton and Nicolle (2013) found that an everyday cold temperature of 5°C had a significant effect on older adults' (+65 years) ability to carry out fine finger tasks and product interactions. Cold temperature can also reduce muscle contraction velocity, which leads to a reduction in the level of force that can be exerted (Holmer 1994, Havenith et al. 1995).



**FIGURE 11.12** Glare caused through bright sunlight reflected off the surface of a product.



*Vibration (whole body)*: when the whole body is brought into vibration via the feet (if standing) or via a seat (if seated) (Griffin 1997, Dul and Weerdmeester 2001). This type of vibration is typically experienced when on transport, and it has the greatest effect on the body's output processes, that is, the ability to maintain hand control (Griffin 1997). In particular, it causes direct mechanical jostling of the hand causing unwanted movement which reduces hand and finger precision; it can also cause unwanted jostling of the display/product being viewed (Griffin 1997).

## Social Environment

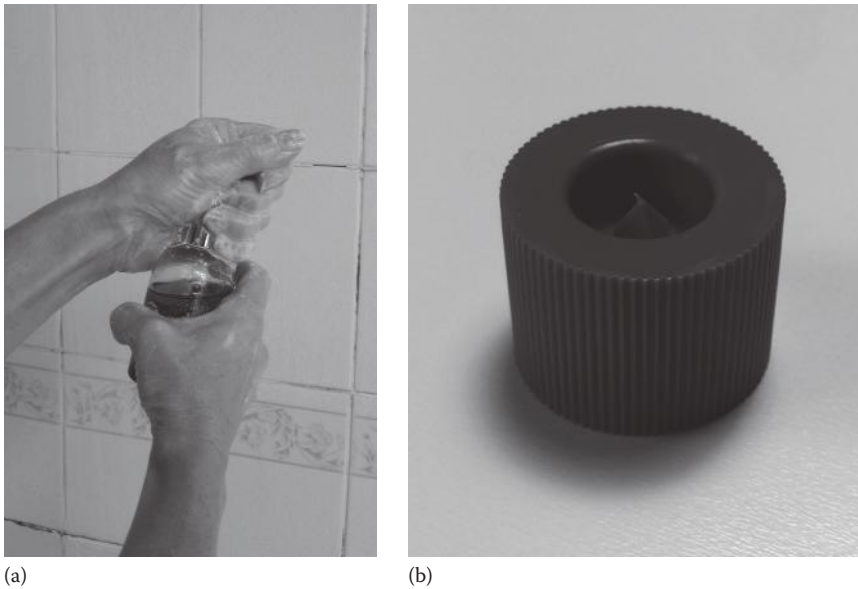
*Group or individual status*: this refers to the presence of others when using a product. Using new products and technology or attending training sessions with other known individual(s) gives people the confidence to interact with and use new products to conduct new tasks, for example attending computer courses, using home computers or surfing the Internet. On the other hand, for some products or tasks, some people may feel put off or embarrassed using them in front of other people. When alone, however, they can be reluctant to try new products, due to a lack of confidence or reassurance from other people. This can result in their constantly using a limited set of products and avoiding more innovative designs, which could be beneficial, for example online shopping. It is important to note, however, that people who form part of a group are not necessarily individuals that will be able to give assistance in using new products or carrying out tasks, but may instead offer company and confidence in both adopting and learning how to use new products or technologies (Goddard and Nicolle 2012).

*Interruptions/distractions*: the completion of certain tasks with products requires a user's attention. Attention can be required for such tasks as remembering a number or a list of instructions or settings, understanding a new interface, navigating through a new menu structure or even writing a text message (Clarkson et al. 2013). Interruptions and distractions can grab or distract a user's attention whilst completing such tasks, making the completion of such tasks either more challenging or even not possible. In a focus group conducted by the authors, an older participant reported that they delayed responding to text messages whilst in busy public places, '... I often wait till I can find somewhere quiet before I reply to text messages so I can concentrate on what to say and do ...'.

## Other Contextual Factors

There are further contextual factors that are not necessarily associated with the environmental context of use, which also need consideration. These factors can be considered as falling under the broader heading of 'context of use', which considers circumstantial factors associated with the user, task, equipment and the environment. For example,

- Heavy bulky clothing (worn for protective reasons or to keep warm) can affect a person's ability to reach and stretch.
- Wet/sweaty/lubricated hands greatly affect dexterity tasks that rely on friction; a higher level of force often needs to be exerted by the users to compensate for such factors. Such conditions are often experienced in kitchens, bathrooms and workshops (Figure 11.13a). Vertical grooved screw tops/caps can help the users obtain a secure grip in such circumstances, as the skin pulps on the hand interlock with the grooves (Figure 11.13b).
- Wearing gloves (e.g. in cold conditions) affects fine finger dexterity; Havenith and Vrijkotte (1993) found that wearing gloves decreased fine finger dexterity by up to 70%. Thus, reducing the number of product interactions that require fine precise hand movements would reduce the impact of the gloves.



**FIGURE 11.13** Contextual factors: (a) wet hands cause reduced friction and (b) grooved cap would enable a more secure grip.

Consideration of such a wide range of contextual factors can become a complicated matter unless there is a structured, integral approach to ensure that it is part of the product design process. The next section provides advice on how to capture this information and to judge when contextual factors as well as product characteristics might result in exclusion.

### IDENTIFYING CONTEXTUAL ISSUES IN THE DESIGN PROCESS

Identifying relevant contextual factors can be achieved in a number of ways; however, the following approach can be used to systematically identify the contextual factors that are most likely to impact on users when interacting with a product.

*Step 1:* Use primary research methods to identify and record the typical contexts of use in which the intended product will be used.

*Step 2:* Conduct a task analysis of using the product (i.e. the steps the user will go through in order to achieve specific goals).

*Step 3:* For each task step, identify what product interaction capabilities will be required to perform the necessary action (e.g. vision and dexterity).

*Step 4:* Determine which of the identified contextual factors will impact on the product interaction capabilities required to use the product.

*Step 5:* Refine the design accordingly.

It is important to carry out user testing at regular points throughout the design process. In particular, test your design with representative users, carrying out the specified tasks, in the identified context(s) of use. Testing will allow you to determine whether you have successfully developed an inclusive product. It will also provide clear design insights as to where refinements are needed.

If we apply this approach to using a mobile phone to make a call, we can see how it can be used to systematically identify where contextual factors are likely to have the greatest impact on product

**TABLE 11.2**  
**Capability Assessment of an Everyday Product**

| Product (Task)                   | Task Step                                     | Capabilities                           |
|----------------------------------|-----------------------------------------------|----------------------------------------|
| Mobile phone (make a phone call) | Unlock phone                                  | Cognition, vision, dexterity           |
|                                  | Enter number/select contact                   | Cognition, vision, dexterity           |
|                                  | Place phone to ear                            | Reach and stretch, dexterity, strength |
|                                  | Wait while phone rings                        | Hearing                                |
|                                  | Hold a conversation                           | Cognition, hearing                     |
|                                  | End call (identify and press call end button) | Vision, dexterity                      |

*Source:* Extract from Elton, E., Generating and translating context capability data to support the implementation of inclusive design within industry, Doctoral thesis, Loughborough University, Loughborough, U.K., 2012.

interaction and how we can obtain a meaningful design insight. Observations and focus groups may reveal that the target users often use their mobile phones to make calls when travelling to tell their relations they are safe and well. From this very brief insight, we can ascertain that calls are often made whilst travelling, whether by car, train or coach. Therefore, the user is likely to be exposed to whole-body vibration, varying levels of illumination and possibly surrounded by other people (noise and distractions). If we then conduct a task analysis of using a mobile phone to make a call, we can determine the capabilities required to complete this task. An extract of such a capability assessment is found in Table 11.2.

From this task analysis, we can see that cognition, vision and dexterity are the key capabilities used. Findings from the user research would suggest that each of these key capabilities are likely to be affected by the identified contextual factors, that is, cognition is likely to be affected due to the presence of other people, vision by the varying levels of ambient illumination, and dexterity is likely to be affected by whole-body vibration. Meaningful design insight can then be gathered from this process in terms of where design modifications are needed to minimise the impact of context.

In this given scenario, the design team should focus on

- Aiding cognition through affordances or by reducing the demands placed on short-term memory, that is, reducing the number of steps involved in making the call, and reducing the memorable information required to unlock the phone. Could the phone be unlocked using an alternative method other than a pin number, for example a finger scanner?
- Enhancing the visual characteristics of the product’s features through increasing their size (stroke-width) and contrast.
- Reducing the number of fine finger precision tasks required to make a phone call, either through shortcut functions or by using alternative methods of interaction to make the call.

This combination of methods allows us to systematically understand the complexity of product interaction and context of use and to gather meaningful design insights, which can help to minimise the impact of context of use on product interaction.

**SUMMARY OF ENVIRONMENTAL CONTEXT OF USE IN INCLUSIVE DESIGN**

Context of use covers a wide and diverse range of factors, which can at first appear overwhelming. However, as detailed, a systematic approach can be employed, which utilises a range of research methods to identify the likely contextual factors that will impact on the design of the product. When designing and developing inclusive products, it is essential that environmental contextual factors are considered due to the impact they can have on users’ capabilities, especially

older users. The following resource is a good place to start to understand the multifaceted impact of context: [www.designingwithpeople.rca.ac.uk](http://www.designingwithpeople.rca.ac.uk) (information on context is displayed under the 'Activities' section).\*

## SUMMARY AND CONCLUDING REMARKS

The inclusive design approach focuses on understanding the needs, preferences and capabilities of people who have some form of impairment and then applying that knowledge to mainstream design. Gaining this understanding can be achieved in a number of ways: through qualitative methods such as interviews and observation, or through quantitative methods such as gathering and applying capability data and design metrics to the design of products, technologies and services. Thus, ergonomics/human factors (i.e. understanding people and applying it to design) plays a significant role in the inclusive design approach.

This chapter has outlined many different reasons for adopting an inclusive design approach, but the most important one is that, putting it simply, 'inclusive design is better design' (Clarkson et al. 2013). Following an inclusive design approach will help you to deliver 'mainstream products and/or services that are accessible to, and usable by, people with the widest range of abilities within the widest range of situations without the need for special adaptation or design' (BS 7000-6 2005). Trying to make a product usable by absolutely everyone is a difficult or impossible task. There will always be people who will require an assistive device or another person to facilitate access, so we must not overlook the possible need for assistive products and technologies. Furthermore, each one of us can experience problems in certain contexts of use, for example using controls when our hands are cold or when we are wearing gloves, viewing a kiosk screen in bright sunlight, driving in difficult conditions, making a telephone call in a noisy environment or when travelling in a foreign country. Vanderheiden (1990) noted that the disabled community refers to those without disabilities as 'TABs', or the Temporarily Able-Bodied, and we could say that this applies to either a person's capability or the contextual situation in which the product is used. Since all of us are ageing and will experience difficulties or impairments as we grow older, if not before, it makes sense to 'design for our future selves' (Benktzon 1993, Coleman and Pullinger 1993, Coleman 2001). Therefore, we encourage designers and developers to consider the needs of a wider range of users and to follow an inclusive design approach as an integral part of their user-centred methods, tools and techniques.

## REFERENCES

- ANEC (Association européenne pour la coordination de la représentation des consommateurs pour la normalisation), the European consumer voice in standardisation. Available at [www.anec.eu](http://www.anec.eu).
- Barrett, J. and Kirk, S. 2000. Running focus groups with elderly and disabled elderly participants. *Applied Ergonomics*, 31(6), 621–629.
- Benktzon, M. 1993. Designing for our future selves: The Swedish experience. *Applied Ergonomics*, 24(1), 19–27.
- British Standards Institution. 2005. BS 7000-6:2005 design management systems. Managing inclusive design. London, U.K.: British Standards Institution.
- Carmeli, E., Patish, H. and Coleman, R. 2003. The aging hand. *Journals of Gerontology Series A: Biological and Medical Sciences*, 58(2), 146–152.
- Cassim, J., Coleman, R., Clarkson, J. and Dong, H. 2007. Why inclusive design? In: R. Coleman, J. Clarkson, H. Dong and J. Cassim (eds.), *Design for Inclusivity*. Aldershot, U.K.: Gower.
- Clarkson, J. 2007. Countering design exclusion – theory and practice. In: R. Coleman, J. Clarkson, H. Dong and J. Cassim (eds.), *Design for Inclusivity*. Aldershot, U.K.: Gower, pp. 165–180.
- Clarkson, J. and Coleman, R. 2010. Editorial: Inclusive design. *Journal of Engineering Design*, 21(2–3), 127–129.

---

\* This resource was created as part of the EPSRC i-design3 project; the resource was created by the Helen Hamlyn Centre for Design at the Royal College of Art and the content on context was developed at Loughborough Design School.

- Clarkson, P.J. et al., *Towards Better Design*, 2010 [computer file]. Colchester, U.K.: UK Data Archive [distributor], May 2012. SN: 6997, doi: 10.5255/UKDA-SN-6997-1.
- Clarkson, P.J., Coleman, R., Hosking, I. and Waller, S. 2013. *Inclusive Design Toolkit*. Cambridge, U.K.: Engineering Design Centre, University of Cambridge. Available at [www.inclusivedesigntoolkit.com/](http://www.inclusivedesigntoolkit.com/).
- Clarkson, P.J. and Keates, S. 2003. User capabilities and product demands. In: *Proceedings of Include 2003*, Helen Hamlyn Research Centre, Royal College of Art, London, U.K., pp. 10:384–10:389.
- Coleman, R. 1994. The case for inclusive design – an overview. In: *Proceedings of the 12th Triennial Congress*, The International Ergonomics Association and The Human Factors Association of Canada, Toronto, Ontario, Canada.
- Coleman, R. 2001. Designing for our future selves. In: W. Preiser and E. Ostroff (eds.), *Universal Design Handbook*. New York: McGraw-Hill, pp. 4.1–4.25.
- Coleman, R. and Pullinger, D.J. 1993. Designing for our future selves. *Applied Ergonomics*, 24(1), 3–4.
- Cook, S.E., Gibb, A.G.F., Richardson, J.H., Walmsley, K. and Bullock, D. 2012. Loughborough University SKInS: Wearable simulations of occupational health – Defining specifications and product development. In: *Proceedings of CIB WO99 International Conference*, Singapore.
- Crossley, L. 2003. Bridging the emotional gap. In: D. McDonagh, P. Hekkert, J. van Erp and D. Gyi (eds.), *Design and Emotion*. Boca Raton, Florida: CRC Press.
- Desrosiers, J., Hébert, R., Bravo, G. and Dutil, E. 1995. The Purdue Pegboard Test: Normative data for people aged 60 and over. *Disability and Rehabilitation*, 17(5), 217–224.
- Diffrient, N., Tilley, A.R. and Bardagjy, J. 1978. *Humanscale. 1/2/3*. Cambridge, MA: MIT Press.
- Diffrient, N., Tilley, A.R. and Bardagjy, J. 1981a. *Humanscale. 4/5/6*. Cambridge, MA: MIT Press.
- Diffrient, N., Tilley, A.R. and Bardagjy, J. 1981b. *Humanscale. 7/8/9*. Cambridge, MA: MIT Press.
- Dong, H., Keates, S. and Clarkson, P.J. 2003. UK and US industrial perspectives on inclusive design. In: *Proceedings of Include 2003*, Helen Hamlyn Research Centre, Royal College of Art, London, U.K.
- Dong, H., Nicolle, C., Brown, R. and Clarkson, J. 2007. Designer-orientated user research methods. In: R. Coleman, J. Clarkson, H. Dong and J. Cassim (eds.), *Design for Inclusivity*. Aldershot, U.K.: Gower, pp. 131–147.
- Dul, J. and Weerdmeester, B. 2001. *Ergonomics for Beginners*, 2nd edn. London, U.K.: Taylor & Francis Group.
- Elton, E. 2012. Generating and translating context capability data to support the implementation of inclusive design within industry. Doctoral thesis. Loughborough University, Loughborough, U.K.
- Elton, E., Johnson, D., Nicolle, C. and Clift, L. 2013. Supporting the development of inclusive products: The effects of everyday ambient illumination levels and contrast on older adults' near visual acuity. *Ergonomics*, 56/5, 803–817.
- Elton, E. and Nicolle, C. 2010. The importance of context in inclusive design. In: *Contemporary Ergonomics 2010, Proceedings of the Annual Conference of the Ergonomics Society*, Keele, U.K.
- Elton, E. and Nicolle, C. 2013. Designing inclusive products for everyday environments: The effects of everyday cold temperatures on older adults' dexterity. *Universal Access in the Information Society (UAIS)*, 12(3), 247–261.
- Elton, E., Nicolle, C. and Mitchell, V. April 2008. Identifying contextual factors in inclusive design. In *Proceedings of Fourth Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT)*, Cambridge, U.K., pp. 14–16.
- Eurostat. 2014. Population structure and ageing, Statistics explained. Available at: [epp.eurostat.ec.europa.eu/statistics\\_explained/index.php/Population\\_structure\\_and\\_ageing](http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_structure_and_ageing) [Accessed 22 November 2014].
- Goddard, N. and Nicolle, C. 2012. What is good design in the eyes of older users? In: P. Langdon, J. Clarkson, P. Robinson, J. Lazar and A. Heylighen (eds.), *Designing Inclusive Systems*. London, U.K.: Springer-Verlag.
- Goodman-Deane, J., Waller, S., Collins, A.-C. and Clarkson J. 2013. Simulating vision loss: What levels of impairment are actually represented? In: M. Anderson (ed.), *Ergonomics and Human Factors 2013*. Cambridge, U.K.: Institute of Ergonomics & Human Factors.
- Goodman-Deane, J., Waller, S., Williams, E., Langdon, P. and Clarkson, P.J. 2011. Estimating exclusion: A tool to help designers. In: *Proceedings of Include 2011*, Royal College of Art, London, U.K., 18–20 April 2011.
- Griffin, M.J. 1997. Vibration and motion. In: G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics*, 2nd edn. New York: John Wiley & Sons Inc, pp. 828–854.
- Grundy, E., Ahlburg, D., Ali, M., Breeze, E. and Sloggett, A. 1999. Research report 94: Disability in Great Britain: Results from the 1996/97 disability follow-up to the family resources survey. London, U.K.: Corporate Document Services.



- Gyi, D.E., Porter, J.M. and Case, K. 2000. Design practice and designing for all. In: *Human Factors and Ergonomics Society Annual Meeting Proceeding; Proceedings 6 – Multiple-Session Symposium*, San Diego, CA, pp. 913–916.
- Havenith, G., Heus, R. and Daanen, H.A.M. 1995. The hand in the cold, performance and risk. *Arctic Medical Research*, 54(2), 37–47.
- Havenith, G. and Vrijkotte, T.G.M. 1993. Effectiveness of personal protective equipment for skin protection while working with pesticides in greenhouses. Part III, Comfort and ergonomics. Report TNO. Soesterberg, the Netherlands: Human Factors Research Institute.
- Henley Centre for Forecasting (now called The Futures Company). 1996. Leisure tracking survey; Family expenditure survey. *Leisure Futures*, Vol. 3, pp. 7–9. London, U.K.: The Futures Company Worldwide.
- Hirvelä, H., Koskela, P. and Laatikainen, L. 1995. Visual acuity and contrast sensitivity in the elderly. *Acta Ophthalmologica Scandinavica*, 73(2), 111–115.
- Hitchcock, D.R., Lockyer, S.E., Cook, S.E. and Quigley, C.L. 2001. Third age usability and safety – an ergonomics contribution to design. *International Journal of Human-Computer Studies*, 55(4), 635–643.
- Holmer, I. 1994. Cold stress part 1: Guidelines for the practitioner. *International Journal of Industrial Ergonomics*, 14(1–2), 139–149.
- Huber, J. and Skidmore, P. 2003. *The New Old*. London, U.K.: DEMOS.
- Hudson, R.B. and Gonyea, J.G. 2012. Baby boomers and the shifting political construction of old age. *The Gerontologist*, 52(2), 272–282.
- International Organization for Standardization (ISO). 1998. BS EN ISO 9241-11:1998 Ergonomics of human-system interaction – part 11: Guidance on usability.
- International Organization for Standardization (ISO). 2006. BS ISO 20282-1:2006 Ease of operation of everyday products – part 1: Design requirements for context of use and user characteristics.
- International Organization for Standardization (ISO). 2010. BS EN ISO 9241-210:2010 Ergonomics of human-system interaction – part 210: Human-centred design for interactive systems.
- Iwarsson, S. et al. 2007. Importance of the home environment for healthy ageing: Conceptual and methodological background of the European project ENABLE-AGE project. *The Gerontologist*, 47(1), 78–84.
- Jackson, G.R., Owsley, C. and McGwin, G., Jr. 1999. Aging and dark adaptation. *Vision Research*, 39(23), 3975–3982.
- Johnson, D., Clarkson, P.J. and Huppert, F. 2009. Capability measurement for inclusive design. *Journal of Engineering Design*, 21(2&3), 275–288.
- Jordan, P. 2000. *Designing Pleasurable Products*. London, U.K.: Taylor & Francis Group.
- Keates, S. and Clarkson, P.J. 2004. *Countering Design Exclusion: An Introduction to Inclusive Design*. London, U.K.: Springer-Verlag.
- Keates, S., Clarkson, P.J. and Robinson, P. 2002. Developing a practical inclusive interface design approach. *Interacting with Computers*, 14(4), 271–299.
- Marshall, R., Case, K., Porter, J.M., Summerskill, S.J., Gyi, D.E., Davis, P.M. and Sims, R.E. 2010. HADRIAN: A virtual approach to design for all. *Journal of Engineering Design*, 21(2&3), 253–273.
- Marshall, R., Cook, S., Mitchell, V., Summerskill, S., Haines, V., Maguire, M., Sims, R., Gyi, D. and Case, K. 2015. Design and evaluation: End users, user datasets and personas. *Applied Ergonomics*, 46(Part B), 311–317, ISSN: 0003-6870. doi: 10.1016/j.apergo.2013.03.008.
- Marshall, R., Summerskill, S.J., Case, K., Gyi, D.E. and Sims, R.E. 2010. Development and evaluation of task based digital human modelling for inclusive design. In: V.G. Duffy (ed.), *Advances in Applied Digital Human Modeling, Proceedings of the Third International Conference on Applied Human Factors and Ergonomics*, Miami, Florida, USA, pp 129–138.
- Maslow, A.H. 1987. *Motivation and Personality*, 3rd edn. Revised by R. Frager et al. New York: Addison-Wesley Educational Publishers Inc.
- Morris, J., Mueller, J. and Jones, M. 2009. Tomorrow's elders with disabilities: What the wireless industry needs to know. *Journal of Engineering Design*, 21(2&3), 131–146.
- Nicolle, C. and Burnett, G.E. 2001. Accessible systems for transport. In: C. Nicolle and J. Abascal (eds.), *Inclusive Design Guidelines for HCI*. London, U.K.: Taylor & Francis Group, pp. 193–207.
- Nicolle, C., Ross, T. and Richardson, S.J. 1993. Identification and grouping of requirements for drivers with special needs. In: *Proceedings of ECART 2 (European Conference on the Advancement of Rehabilitation Technology)*, Stockholm, Sweden, 26–28 May, 1993. Session 3.1.
- North Carolina State University, Centre for Universal Design. 1997. Seven principles of universal design. Available at: [www.ncsu.edu/ncsu/design/cud/](http://www.ncsu.edu/ncsu/design/cud/).



- Office for National Statistics (ONS). 2011. 2010-based national population projections. Available at: [www.ons.gov.uk/ons/rel/npp/national-population-projections/2010-based-projections/rep-2010-based-npp-results-summary.html](http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2010-based-projections/rep-2010-based-npp-results-summary.html). [Accessed 22 November 2014].
- Office for National Statistics (ONS). 2012. Income of retired households, 1977-2010/11. Available at: [www.ons.gov.uk/ons/rel/household-income/income-of-retired-households/1977-2010-11/rep-income-of-retired-households-1977-2010-11.html](http://www.ons.gov.uk/ons/rel/household-income/income-of-retired-households/1977-2010-11/rep-income-of-retired-households-1977-2010-11.html).
- Ostroff, E. 2001. Universal design: The new paradigm. In: W.F.E. Preiser and E. Ostroff (eds.), *Universal Design Handbook*. New York: McGraw-Hill, pp. 1.3–1.12.
- Persad, U., Langdon, P.M. and Clarkson, P.J. 2006. Inclusive design evaluation and the capability-demand relationship. In: P.J. Clarkson, P. Langdon and P. Robinson (eds.), *Designing Accessible Technology*. London, U.K.: Springer-Verlag, pp. 177–183.
- Pheasant, S. and Haslegrave, C.M. 2006. *Bodyspace: Anthropometry, Ergonomics and the Design of Work*, 3rd edn. London: Taylor & Francis Group.
- Poulson, D., Ashby, M. and Richardson, S.J. (eds.) 1996. *USERfit: A Practical Handbook on User Centred Design for Assistive Technology*. Commission of the European Communities. Brussels, Luxembourg: ECSC-EC-EAEC. Main sections available on the EDeAN website at [www.edean.org/index.php?card\\_id=47&filters=f27&row=6&collect=47&query](http://www.edean.org/index.php?card_id=47&filters=f27&row=6&collect=47&query). [Accessed 22 November 2014].
- Research Institute for Consumer Affairs (Rica). 2001. Inclusive design – products that are easy for everybody to use. Stratford upon Avon, England: Disability Rights Commission (now part of the Equality and Human Rights Commission).
- Richardson, S.J., Poulson, D.F. and Nicolle, C. 1993. User requirements capture for adaptable smarter home technology. In: E. Ballabio, I. Placencia-Porrero and R. Puig de la Bellacasa (eds.), *Rehabilitation Technology*. Amsterdam, the Netherlands: IOS Press, pp. 244–248.
- Royal National Institute of Blind People (RNIB). 2006. See it Right: Making information accessible to people with sight problems. London, U.K.: Royal National Institute of Blind People.
- Sainio, P. et al. 2006. Self-reported and test-based mobility limitations in a representative sample of Finns aged 30+. *Scandinavian Journal of Public Health*, 34(4), 378–386.
- Sanders, M.S. and McCormick, E.J. 1993. *Human Factors in Engineering and Design*, 7th edn. New York: McGraw-Hill Inc.
- Sims, R.E., Maguire, M.C., Nicolle, C., Marshall, R., Lawton, C. Peace, S. and Percival, J. 2012. Older people's experiences of their kitchens: 2000 to 2010. *Housing, Care and Support*, 15(1), 6–15.
- Smith, S., Norris, B. and Peebles, L. 2000. *Older Adultdata: The Handbook of Measurements and Capabilities of the Older Adult – Data for Design Safety*. London, U.K.: Department for Trade and Industry.
- Summerskill, S.J. et al. 2009. Validation of the HADRIAN system with a train station design case study. In: P.D. Bust (ed.), *Contemporary Ergonomics 2009, Proceedings of the Annual Conference of the Ergonomics Society*. London, U.K.: Taylor & Francis Group, pp. 70–78.
- Tenneti, R., Johnson, D., Goldenberg, L., Parker, R.A. and Huppert, F.A. 2012. Towards a capabilities database to inform inclusive design: Experimental investigation of effective survey-based predictors of human-product interaction. *Applied Ergonomics*, 43(4), 713–726.
- Trombly, C.A. and Scott, A.D. 1989. Evaluation of motor control. In: C.A. Trombly (ed.), *Occupational Therapy for Physical Dysfunction*, 3rd edn. Baltimore, MD: Williams & Wilkins.
- United Kingdom Government Equalities Office. 2010. Equality Act. [www.gov.uk/equality-act-2010-guidance](http://www.gov.uk/equality-act-2010-guidance). [Accessed 22 November 2014].
- United Nations. 2008. Convention on the rights of persons with disabilities. [www.un.org/disabilities/convention/conventionfull.shtml](http://www.un.org/disabilities/convention/conventionfull.shtml). [Accessed 22 November 2014].
- United States Census Bureau. 2011. The older population: 2010. [www.census.gov/prod/cen2010/briefs/c2010br-09.pdf](http://www.census.gov/prod/cen2010/briefs/c2010br-09.pdf). [Accessed 22 November 2014].
- United States Department of Justice. 1990. Americans with Disabilities Act. [www.ada.gov/pubs/adastatute08.htm](http://www.ada.gov/pubs/adastatute08.htm). [Accessed 22 November 2014].
- Vanderheiden, G.C. 1990. Thirty-something (million): Should they be exceptions? *Human Factors*, 32(4), 242–250.
- Yelding, D. and Cassim, J. 2007. Practicalities of working with users. In: R. Coleman, J. Clarkson, H. Dong and J. Cassim (eds.), *Design for Inclusivity*. Aldershot, U.K.: Gower, pp. 149–163.



---

# 12 Analysis and Design of Warnings in the Workplace

*Christopher B. Mayhorn, Michael S. Wogalter  
and Kenneth R. Laughery*

## CONTENTS

|                                                                              |     |
|------------------------------------------------------------------------------|-----|
| Introduction.....                                                            | 331 |
| Defining the Work Setting.....                                               | 332 |
| Hazard Hierarchy .....                                                       | 332 |
| Purpose of Warnings .....                                                    | 333 |
| Recognised Challenges of Warning Design and Evaluation .....                 | 333 |
| Warning Systems.....                                                         | 333 |
| Components.....                                                              | 334 |
| Work Settings .....                                                          | 334 |
| C-HIP Model.....                                                             | 336 |
| Source.....                                                                  | 338 |
| Channel .....                                                                | 338 |
| Delivery .....                                                               | 339 |
| Environmental Stimuli .....                                                  | 339 |
| Receiver.....                                                                | 340 |
| Attention Switch.....                                                        | 340 |
| Attention Maintenance .....                                                  | 342 |
| Comprehension and Memory .....                                               | 344 |
| Beliefs and Attitudes .....                                                  | 348 |
| Motivation .....                                                             | 350 |
| Behaviour .....                                                              | 350 |
| Receiver Variables .....                                                     | 351 |
| A Look into the Future: Twenty-First-Century Inclusive Warning Systems ..... | 351 |
| Summary and Utility of the C-HIP Model.....                                  | 352 |
| References.....                                                              | 353 |

## INTRODUCTION

Work is an integral component of human society. Much, if not most, of the world population spends a large portion of its waking hours in the workplace setting (U.S. Bureau of Labor Statistics, 2012). Not surprisingly, job satisfaction and productivity are linked to life satisfaction (Keon and McDonald, 1982). Given this connection, one main objective of the ergonomics/human factors (E/HF) discipline includes enhancing workplace productivity and safety (Sanders and McCormick, 1993). One approach to achieving this objective includes the development and strategic use of effective warnings as a means of communicating safety-related information to workers. This chapter focuses on the factors that influence warning effectiveness within the workplace using the

communication-human information processing (C-HIP) model. Not only does this model serve as an effective means to review the warnings literature, but it also can be used as a predictive tool to understand why certain warnings fail.

## DEFINING THE WORK SETTING

Any work setting in which we may find hazards and/or warnings varies considerably based on a number of factors such as geographic location, the type of work being performed and the characteristics of the workforce, to name but a few. The workplace of a factory assembly line worker may greatly differ from the workplace of a university professor; yet some commonalities exist to inform E/HF practitioners how manipulations of warning characteristics can generalise to many workplace contexts. To understand the complex interactions that occur between people, equipment, tasks and environments within specific settings, a standard practice in the field of E/HF is to apply a systems approach (Helander, 1997). Thus, a work system environment includes a variety of elements such as the working person, the work task, operating resources and the work environment (Luczak et al., 2006 and see Chapter 1 of this book). A working person can be described using a variety of dimensions such as the physical capability to perform work, including functional limitations as well as the psychological dimension of willingness to perform work. In the case of the assembly line worker, it might be important to note physical limitations that could influence work, such as a lower back injury that prevents the lifting of objects weighing in excess of 30 lb. Likewise, the university professor might be differentially motivated to achieve one task (e.g. writing a journal article) instead of another (e.g. attending a faculty meeting).

Work tasks can be defined as the step-by-step procedure for fulfilling an objective, whereas operating resources include the equipment needed to perform the work. In the case of the assembly line worker, equipment such as a drill press and other tools might be used in the task of machining automobile parts, whereas the university professor might operate a scanner and desktop computer to document travel receipts. Finally, the work environment includes the social and cultural factors as well as the physical attributes of the environment where work is performed. Differences in the workplaces of the exemplar assembly line worker and professor might include the amount of lighting available as well as the presence of excessive noise from equipment and other personnel. The social differences in a workplace environment might be seen in, for example, institutional differences in the *safety culture* demonstrated by groups of workers that might encompass basic assumptions, attitudes or values regarding organisational safety concerning potential hazards in the workplace (Marquardt et al., 2012).

## HAZARD HIERARCHY

Different degrees of safety culture can exist between organisations such that one could have a very low, almost fatalistic level (e.g. ‘accidents are bound to happen on the job’) and another has a high level (‘we are proactive in solving safety problems before they arise’; see Parker et al., 2006). Consistent with the more proactive approach to hazard mitigation, a well-accepted general hierarchy of hazard control is often associated with efforts to reduce hazards (Laughery and Wogalter, 2006; Lenorovitz et al., in press). It includes several fundamental strategies or approaches that can be employed to limit occupational risks. The first and best strategy for controlling hazards is to design them out – to eliminate them. For example, if a traditional manufacturing process requires the use of toxic chemicals to create a product, an employer might seek alternate manufacturing procedures that utilise non-toxic chemicals. Sometimes the hazard cannot be completely eliminated without ruining the functionality of the product or otherwise impacting quality, so a second-level strategy of hazard control, guarding, should be considered. With guarding, the hazard remains present, but there is a barrier to separate the worker from the hazard to prevent harm. A guarding method might employ a better way to keep the toxic chemicals from coming in contact with workers. For instance,

the manufacturing steps that require the use of toxic chemicals might be conducted when workers are isolated in another room; alternately, if workers have to be present, they might be outfitted with protective clothing.

Warnings are the third strategy of hazard control. Only when the strategies of designing out or guarding are not feasible or practical should warnings be the chosen method of controlling hazards. Design and guarding methods should be considered, tested and used if they reduce hazards and do not dramatically hurt functional utility. Also, even with design and guarding changed, warnings might still be needed if the workplace hazards remain after design and guarding considerations have been made and any enactments of them have taken place. In general, good design and guarding are better methods of hazard control than warnings. Indeed, warnings are properly viewed as a supplement, not a substitute, to other approaches to safety (Lehto and Salvendy, 1995). Therefore, if the goal is to protect workers from harm, when the hazard has not been controlled by design/guarding, then the warning system needs to be designed to maximise its effectiveness so as to influence people's perceptions, cognition and behaviour.

### **PURPOSE OF WARNINGS**

Warnings in the workplace serve three main purposes (Conzola and Wogalter, 2001). First, warnings are used to improve safety by reducing the likelihood of workplace accidents that might result in death, personal injury or property damage. Second, they are used to communicate important safety-related information to a target audience such as workers or others (e.g. visitors) present in the workplace. In general, warnings should include a description of the hazard, instructions to avoid the hazard and the consequences that might occur as a result of not complying with the warning (Rogers et al., 2000). Finally, warnings are used to promote safe behaviour and reduce unsafe behaviour. For example, warnings might serve to remind employees of their previous safety training where they were instructed to don personal protective equipment such as earplugs or face shields (Leonard et al., 1999).

### **RECOGNISED CHALLENGES OF WARNING DESIGN AND EVALUATION**

While the need for effective workplace warnings should be apparent, a number of challenges may slow or even prevent the deployment of warnings that can protect worker safety (Laughery, 2006). For instance, as international trade grows, the increasing diversity of workers will make it difficult to design a warning that addresses the needs of all individual workers. Consider a workplace where people speak multiple languages, have different literacy levels and have different cultural values. It is almost the Herculean task of a warning designer to develop a warning that identifies the hazard and promotes safe behaviour by informing workers how to avoid being injured or killed. Once such a prototype warning is developed, it should undergo an iterative process that includes rigorous evaluation and redesign whereby observed shortcomings in earlier warning design are corrected as the design is updated (Mayhorn and Goldsworthy, 2009).

### **WARNING SYSTEMS**

When asked to consider the concept of a warning, many individuals take a very narrow view and believe that safety information is transmitted solely as a static sign or a portion of a label (Laughery and Wogalter, 1997; Wogalter and Mayhorn, 2005). In Figure 12.1, a label is attached to a piece of industrial equipment that poses a heat hazard that might result in a fire or a thermal burn. While this component is certainly important, it is often necessary to broaden the transmission of safety information to include several components in the form of a warning system that utilises a variety of media and messages (Laughery and Hammond, 1999).



**FIGURE 12.1** Warning on a heat gun. (From Richard M. Hansen & Associates, Inc., Lombard, IL, File No. 98-446.0244, 12 May 2000.)

## COMPONENTS

Warning systems usually contain multiple components. For example, consider the warning system designed to assist mechanics in the matching of automobile tyres and rims where the consequences of a mismatch could result in explosive tyre decompression leading to the death of a motorist. Components of this tyre warning system may include the raised lettering on the side wall of tyres, tread labels on new tyres, stickers or stamping on the rim, safety-related wall posters in shops where tyres are mounted, statements in rim and tyre catalogues and manuals, verbiage in documentation accompanying sales receipts of tyres and rims, reminder pop-up messages in point of sales electronic workstations and verbal statements from employers and other employees. It should be noted that these components may not be identical in terms of content or purpose. For instance, some components may contain minimal content and be intended to capture attention so that it functions to direct a person to another component that includes more comprehensive information. An example is the statement ‘See the Michelin Fitment Guide’, which might be included on a handout to remind tyre professionals that they can accurately match the tyre size that is appropriate to a particular rim by accessing this comprehensive manual.

Because workplaces differ, some components of warning systems may be unique to certain settings. For instance, different components may be intended for different target audiences. Within agricultural settings, warning components that accompany pesticides might include printed on-product labels, verbiage in advertisements about the product, verbal statements from the salesperson to the buyer as well as material safety data sheets (MSDSs) which include sophisticated chemical information regarding the product. In this instance, the on-product warning is intended for everyone who comes in contact with the pesticide including a farm worker, whereas the MSDS is directed to a safety professional working for the employer. Thus, informational content will differ due to different purposes of the information and the characteristics of the target audiences.

## WORK SETTINGS

In the workplace as opposed to home or other non-workplace environments, there is control over what people do. Employees are different than non-employee personnel because employers have



contractual arrangements that can affect what they can do. Employers in the workplace can control many aspects, including providing safety training. They can potentially train workers about everything they do. In the United States, the Occupational Safety and Health Administration (OSHA) is a government agency tasked with enforcing workplace safety. Although OSHA requirements mandate the need for regular safety meetings, it is unclear how the quality of those training sessions can be assessed. Some employers undoubtedly have very good training, but smaller organisations that employ fewer workers may not have the time or resources to do everything (paperwork included) effectively. Manufacturers need to expect that employers may not be able to take much effort to train their employees about every hazard to which they may be exposed. Product and equipment manufacturers need to provide to businesses and government agencies, etc. materials that make it easy to train workers to be productive, satisfied and safe.

Based on this discussion, it is clear that a working person often relies on others as resources to provide appropriate components within a warning system. As the concept of *cascading responsibility* described by Williams et al. (2006) suggests, multiple parties are involved in the dissemination of safety information. In the case of a tree removal specialist, an employee might be tasked with using a woodchipper to clear debris, and he or she is dependent on a number of entities to provide safe job-related instruction. Employers/supervisors should provide adequate training to promote safety in the field. Such training could include a formalised set of courses, basic *hands-on* training, regular safety meetings or the provision of written policy documents that describe operational procedures. Moreover, these individuals are charged with direct supervision of an employee such that they can monitor safety performance and intervene should the employee deviate from safety protocols. In the United States, employers are responsible for workers' safety, and employees are prevented from suing employers for negligence for almost any reason because federal government laws – workman's compensation laws – prevent lawsuits and have a different method of compensation.

Because employees and employers interact with equipment such as the aforementioned woodchipper, they are dependent on the manufacturer of this device to provide instructions for safe use. Manufacturers know or should know more than other parties about the potential hazards accompanying the use of their product, so they are tasked with employing the hazard hierarchy during product design. If alternate designs cannot be used to eliminate hazards, manufacturers must include appropriate safeguards in their design and disseminate safety information and instructions. Such information might include warnings in a variety of components such as MSDs, informational inserts/pamphlets, owner's manuals or on-product labels. For example the exterior of a woodchipper that was literally covered with stickers included the following safety text:

DO NOT attempt to operate the machine without proper training and becoming very familiar with the operator's manual. The hydraulic feed wheels are designed to pull wood into the chipper. They do not know the difference between a hand and wood. If a guard is removed, it must be replaced or severe injury can result. The cutter-wheel coasts for several minutes after the power is shut down. DO NOT attempt any maintenance while the wheel is turning. NEVER open or close the cutterwheel cover while the disc is still turning.

When purchasing equipment for use in the workplace, most employers do not deal directly with manufacturers. Instead, employers may interact with intermediaries such as distributors or retailers. For instance, the woodchipper manufacturer may sell its product to distributors who then employ retailers to sell their products to the employer. In such a situation, the manufacturer must be certain that safety information is passed to the distributor who in turn passes the information to the retailer who communicates directly with the employer/buyer. If at any point in the chain of commerce safety instructions are lost, the end user (employee or subcontractor) might be placed at the risk of injury or death. If a manufacturer then dutifully provides the appropriate on-product warning labels but a distributor or retail entity decides to paint the exterior of the woodchipper for aesthetic reasons, safety information may be covered, and end users may not be informed about the hazard.

Unlike the case with employers mentioned earlier, injured users in the United States can sue the distributors, retailers and manufacturers of the equipment for irresponsible design, guarding and warning. So potentially, any of the entities in the line of commerce such as product and equipment manufacturers, retailers and distributors can be found negligent for failing to provide necessary warnings and instructions to the employer and the employee.

Given the potentially lethal consequences that occur when a hazard is not adequately communicated to workers, the design and evaluation of effective warnings are critical. As the contents of this chapter will demonstrate, E/HF professionals face a number of challenges in creating and deploying warnings. Not only must they consider relatively mundane issues such as where to place warnings on products or how to disseminate safety education materials but they must also determine how information should be presented in the particular context of a workplace to multiple stakeholders who vary considerably. Thus, it is necessary to understand how workers process warning information based on fundamental cognitive principles as a means to addressing the real-world problem of safety in the workplace.

## C-HIP MODEL

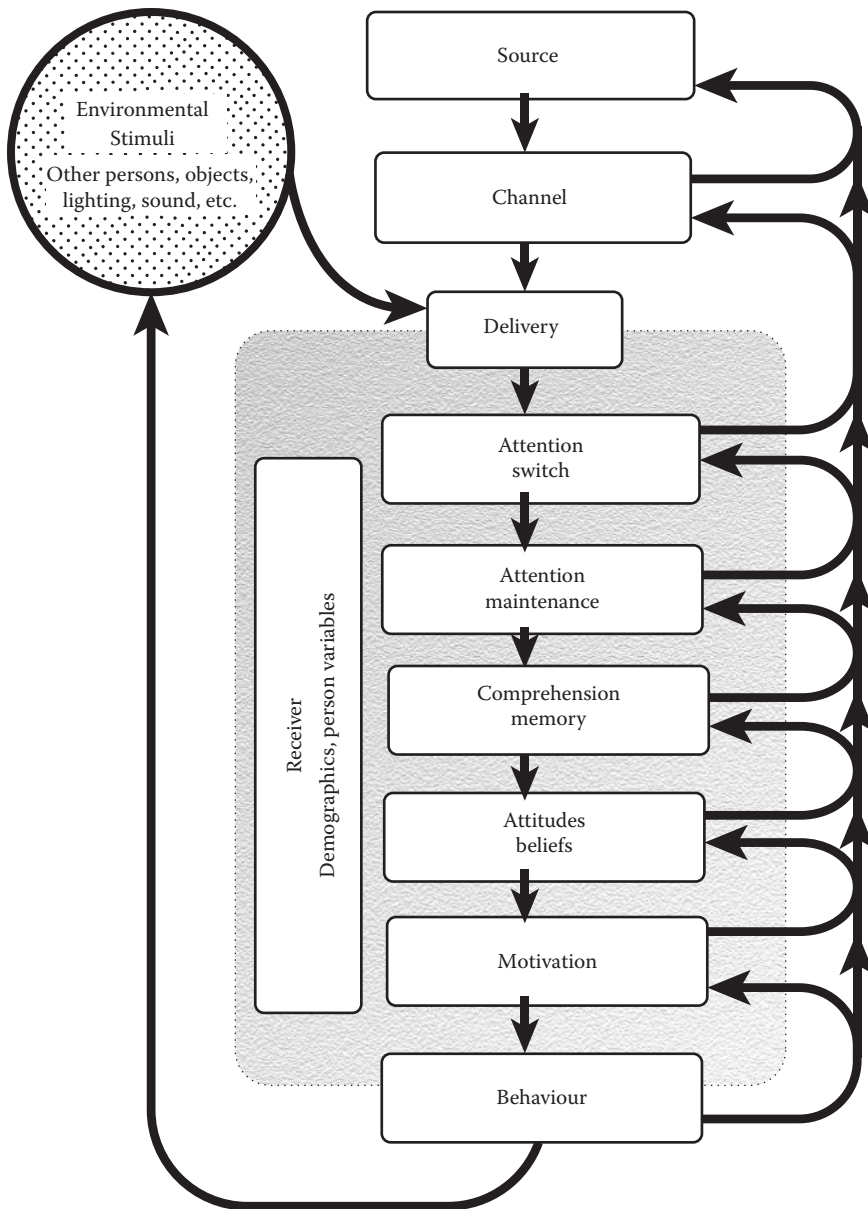
In this section, the C-HIP model is presented to serve as an organising framework for reviewing some of the major concepts and findings regarding factors that influence warning system effectiveness (Wogalter, 2006b). Specifically, this chapter reviews research of some of the influential factors found at each stage. After going through the stages of the model, another benefit of the C-HIP model is described, namely, it can serve as an investigative tool for helping determine why a warning failed to be effective.

The C-HIP model has two major sections, each with several component stages. A representation of the model can be seen in Figure 12.2. The first section of the framework employs the basic stages of a simple communication model. Here the model focuses on a warning message being sent from one entity (source) through some channel(s) to another (receiver).

The second major section of the model focuses on the receiver and how people internally process information. This section interfaces with the first through effective delivery of the warning to individuals who are part of the target audience. When warning information is delivered to the receiver, processing may be initiated and, if it is not blocked in some way, will continue through several stages: from attention switch, attention maintenance, comprehension and memory, beliefs and attitudes, motivation and possibly ending in behaviour. Similar information processing models have been discussed by others (Lehto and Miller, 1986; Rogers et al., 2000).

The C-HIP model is both a stage model and a process model. The model is useful in describing a general sequencing of stages and the effects warning information might have as it is processed. If information is successfully processed at a given stage, the information *flows through* to the next stage. If processing at a stage is unsuccessful, the flow of information will not reach to the next stage. If a person does not initially notice or attend to a warning, then processing of the warning goes no further. However, even if a warning is noticed and attended to, the individual may not understand it, and as a consequence, no additional processing occurs beyond that point. Even if the message is understood, it still might not be believed, thereby causing a blockage to occur at this point. If the person believes the safety message, then low motivation (to carry out the warning's instructed behaviour) could be a cause for a blockage. If all of the stages are successful, the warning process ends in safety behaviour (compliance) attributable to the warning information.

Although the model tends to emphasise a linear sequence from source to behaviour, there are feedback loops from later stages in the process which can impact earlier stages of processing as illustrated on the right side of Figure 12.2. For example, when a warning stimulus becomes habituated from repeated exposures over time, less attention is given to it on subsequent occasions. A more specific example could be given in terms of prescription pharmaceuticals (Guchelaar et al., 2005). If a new hazard is added to a warning, a pharmacist may not notice it if he or she had previously prescribed and read the previous warning version in the past. Here, memory and experience affect



**FIGURE 12.2** C-HIP model.

an earlier attention stage. A second example of feedback effects concerns the influence of beliefs on attention. Some individuals may not believe that a given product is hazardous and, as a result, not think about looking for a warning. Thus, if a pharmacist or other health-care professional believes that a commonly prescribed analgesic can cause no harm, they will be less likely to read a warning about newly found drug interactions (Russ et al., 2012). Thus, a later stage, beliefs and attitudes, affects an earlier stage of attention.

In the following sections, factors affecting each stage of the C-HIP model are described. The first three sections concern the communication features of the C-HIP from the source via some channel(s) to the receiver. Subsequent sections concern analysis of information processing factors that are internal to the receiver.

## SOURCE

The source is the initial transmitter of the warning information. The source can be a person (e.g. supervisor, co-worker) or an organisation (e.g. company, government). With respect to the workplace, warning information often does not originate from the employer; instead, the employer may post pre-prepared safety statements/instructions provided by a third party. Because employees often interact with equipment, this third-party source is frequently the manufacturer of workplace machinery (although in cases of imported products, the importer/distributor in the United States may be held responsible) (McGrath, 2011). One critical role that the source assumes is to determine if there are hazards present that necessitate a warning. Such a determination requires some form of hazard analysis (Frantz et al., 1999; Young et al., 2006). If a hazard is identified, the source must first determine if there are better methods of controlling it than the use of warnings, such as eliminating or designing out the hazard or guarding against it using design and engineering procedures (for a comprehensive review, see Laughery and Wogalter, 2006).

There are other considerations such as the specific characteristics of the equipment involved. Some products are inherently more dangerous than others. For instance, a manufacturer of a nail gun will have a different role to play than a manufacturer of a welding helmet. Although even safety-related products such as welding helmets can also have hazards, it remains the responsibility of the manufacturer to mitigate potential risks, which might include the use of warnings.

If the need for a warning exists, then the source (generally the manufacturer) needs to determine how workers should be warned, for example, what channel(s) to use (see later section) and the warning's intrinsic characteristics. In addition, the perceived characteristics of the source can influence people's beliefs, credibility and relevance (Wogalter et al., 1999d; Cox and Wogalter, 2006). Information from a reliable expert source is usually given greater credibility. It is generally assumed that the manufacturer is expert with regard to the product they produce. It is expected that they know or seek to learn about hazards and keep them at bay. That is their role. If the source does not carry out its role satisfactorily, persons can be injured, and in some cases, depending on country and legal jurisdiction, the manufacturer can be sued, fined and the product recalled.

## CHANNEL

The channel is the medium and modality in which information is transmitted from the source to one or more receivers. Workplace warnings can be transmitted in many ways. Warnings can be presented in labels directly on equipment, on shipping containers, in user manuals, in package inserts, on posters/placards, in brochures and as part of audio–video presentations in various media such as face-to-face safety meetings or via the Internet. Most commonly, warnings use the visual (text and symbols) and auditory (alarms and voice) modalities as opposed to the other senses. There are exceptions, for example, an odour added to industrial glue to signal the olfactory sense thereby reminding users to ventilate the work area (Hattem and Letho, 1995), and the rough vibration of a product that is not mechanically functioning well which provides tactual, kinaesthetic and haptic sensation (Mazis and Morris, 1999; Cohen et al., 2006).

There are two dimensions of the channel. The first concerns the media in which the information is embedded (e.g. label, DVD video). The second dimension is the sensory modality of the receiver (visual, auditory). Some media involve one modality (e.g. product manual involves the visual sense) and others involve two modalities (e.g. DVD or WWW videos often have both visual and auditory). Visual presentation can be comprised of both or either text and symbols. Auditory presentation can be nonverbal (noise, beeps, buzzers) and verbal (voice/speech) sounds. For example, traditional smoke alarms or carbon monoxide detectors produce nonverbal signals.

Multimodal warnings may be more effective in promoting safety behaviour than warnings that utilise a single modality because they provide redundancy (Baldwin et al., 2012). If an individual is not watching a visual display, he/she can still hear it (Wogalter and Young, 1991; Barlow and

Wogalter, 1993). If the individual is blind or deaf, the information is available in the other modality. In addition, if an individual sees and hears warning information, there is a greater likelihood that the message will be delivered to otherwise vulnerable receivers (e.g. both deaf and blind persons will be satisfied, and persons overloaded in one modality could receive it in another modality). Also there is well-supported theory in cognitive psychology and education that multimodal presentation enhances learning because the information is richer and may link to greater or better internal representational nodes (Paivio, 1971; Clark and Paivio, 1991).

Longer, more complex messages may be better presented visually because reading language is generally faster and allows easier review and re-review of the material. However, shorter, less complex messages presented aurally can be more effective than the same messages presented visually. Also, the presentation of an auditory signal is generally better for switching attention (a stage described earlier). An implication from this analysis is that a short auditory warning pointing to more detailed information accessible elsewhere would be beneficial for capturing attention as well as enabling the processing of longer and more complex information.

## DELIVERY

While the source may try to disseminate warnings in one or more channels, the warnings might not reach some of the targets at risk (Williamson, 2006). Delivery refers to the point of reception where a warning arrives at the receiver. To emphasise its importance, it is shown as a separate stage in the current C-HIP model shown in Figure 12.2. A warning that a person sees or hears is a warning that has been delivered. A safety video that is produced by a tool manufacturer but never reaches the workplace would constitute delivery failure. The reason for the failure to deliver the warning to targeted individuals can be multiple. The video may be sitting in bulk boxes in a warehouse and not have been distributed because the manufacturer's ordering procedure requires that an employee must actively request the information. Or the distribution could be haphazard, reaching some intended persons and not others. But even if individuals receive the video, they may not receive the needed information. They may not have the time or playback equipment to see it. Of course, even if the person does see the video, it may not include the necessary warning. Thus, it may be necessary to distribute warning information in multiple ways to reach receivers at risk.

## ENVIRONMENTAL STIMULI

Besides a given workplace warning, other stimuli are almost always simultaneously present. They may be other warnings or a wide assortment of non-warning stimuli. These stimuli compete with the warning for the worker's attention (described further in text). With respect to a given warning, these other stimuli may be described as *noise* that could potentially interfere with warning processing. Several examples can illustrate. A co-worker's cellular phone ringing just when an individual begins to examine a warning may cause distraction and lead to the warning not being fully read. Likewise an abraded, scratched warning from considerable use and environmental exposure could prevent a newly hired worker from reading the warning on the pneumatic nail gun illustrated in Figure 12.3. The environment can have other effects. The illumination can be too dim to read the warning. In such cases of distraction or legibility, another warning of greater salience could have the capability to attract and hold a person's focus.

Environmental influences can also include other people. Awareness of what others are doing in the local environment and elsewhere can affect warning compliance both positively and negatively (deTurk et al., 1999; Olson et al., 2009). Seeing co-workers wearing hard hats at a construction site suggests it is proper behaviour to wear them. But instances of seeing others not wearing goggles, gloves or other needed protective equipment while using a hazardous tool





**FIGURE 12.3** Pneumatic nail gun warning.

can suggest that such protection is not needed, even though the other safety information such as a product warning states that it is required (Wogalter et al., 1989).

## RECEIVER

The receiver is the person(s) or target audience to whom the warning is directed. In the context of the workplace, receivers can be the working person as well as supervisors and others (e.g. visitors) who are exposed to the hazards. For a warning to effectively communicate information and influence behaviour, the warning must first be delivered. Then, attention must be switched to it and maintained long enough for the receiver to extract the necessary information. Next, the warning must be understood and must concur with the receiver's existing beliefs and attitudes. Finally, the warning must motivate the receiver to perform the directed behaviour. The next several sections are organised around these stages of information processing.

## Attention Switch

An effective warning must initially attract attention. To do so, it needs to be sufficiently salient (conspicuous or prominent). Warnings typically have to compete with other stimuli in the environment for attention. Several design factors influence how well warnings may compete for attention (see Wogalter and Leonard, 1999; Wogalter and Vigilante, 2006).

For visual presentation of information, larger is generally better. Increasing the overall size of the warning, its print size and contrast generally facilitate warning conspicuousness. Context also plays an important role. It is not just the absolute size of the warning but also its size relative to other displayed information that matters. Consider a workplace with multiple safety signs depicted in Figure 12.4. Here, the warning text for the signs *Hard Hat Required* and *Eye Protection Area* may be in direct competition with one another.

Colour is an important attribute that can facilitate attracting attention (Laughery et al., 1993b; Bzostek and Wogalter, 1999). While there are potential problems with using colour as the only method of conspicuousness, such as colour blindness in some individuals, colour is a frequently used design component to attract attention. Figure 12.5 is an example of the ANSI Z535 (2002) standard





FIGURE 12.4 Multiple warning signs in a workplace.



FIGURE 12.5 Illustration of the alert symbol in a signal word panel.

that uses colour (red in this example) as one of several components of the signal word panel to attract attention. Other design components in the ANSI Z535 signal word panel include an alert symbol, the triangle/exclamation point and one of three hazard connoting signal words (DANGER, WARNING and CAUTION). Context again can play a role with respect to colour as a salient feature. A red warning on a product label located on a red tool will have less salience than the same warning conveyed using a different colour. The colour should be distinctive in the environment where it is placed.

Symbols can also be useful for capturing attention. One example (depicted with the DANGER signal word panel in Figure 12.5) already mentioned is the alert symbol represented as a triangle enclosing an exclamation point (Laughery et al., 1993b; Bzostek and Wogalter, 1999). This symbol serves as a general alert. Bzostek and Wogalter (1999) found results showing people were faster in locating a warning when it was accompanied by an icon. Other kinds of symbols may be used to convey more specific information. This latter purpose is discussed in the later comprehension section, but the point here is that a graphic configuration can also benefit the attention switch stage.

Warnings located proximal to the hazard, both temporally and physically, generally increase the likelihood of attention switch (Frantz and Rhoades, 1993; Wogalter et al., 1995). Warnings should be located to maximise the chance that they will be encountered. This aids in delivery. For instance, a warning about carbon monoxide (CO) hazards on a gas-powered electrical generator is more likely

to be effective than one located in a separate, sometimes displaced (e.g. in a file or possibly lost or never received), product manual (Wogalter et al., 1998c; Mehlenbacher et al., 2002). Generally, placement directly on the product or its primary container is preferred, particularly if the product is potentially highly dangerous (Wogalter et al., 1991, 1995). There may be exceptions to the proximity rule, such as where the warning is presented too close in location and/or time to the hazard, and the individual sees or hears it too late to avoid the hazard.

Repeated long-term exposure to a warning may result in a loss of its ability to evoke an attention switch at later times (Thorley et al., 2001; Kim and Wogalter, 2009). This process or state of habituation can eventually occur even with well-designed warnings; however, better designed warnings with salient features can slow the habituation process. As discussed in Wogalter and Mayhorn (2005), one of the benefits of technology-based warnings is that a warning's appearance can be changed to reinvigorate attention switch previously lost due to habituation.

Work-related tasks the individual may be performing and other stimuli in the environment may absorb attention and may compete with the warning for attention capture (Wogalter and Usher, 1999). While previous examples have focused on the visual modality, auditory perception is important as well. Consider the warehouse worker who is engaged in some task within a noisy environment where heavy equipment such as forklifts are in operation. Such a situation may present a safety hazard because auditory back-up alarms meant to alert workers to an approaching forklift must compete with other stimuli such as engine noise, talking co-workers or the presence of hearing protection devices that may detract from a worker's attention (Robinson and Casali, 1995). Thus, the warning should have characteristics to make it highly salient in context.

### Attention Maintenance

Individuals may notice the presence of a warning but not stop to examine it. A warning that is noticed but fails to maintain attention long enough for its content to be understood may be of very little direct value. Attention must be maintained on the message for some length of time to extract meaning from the material. During this process, the information is encoded or assimilated with existing knowledge in memory.

With brief text or symbols, the warning message may be grasped very quickly, sometimes as fast as a glance. For longer more complex warnings, attention must be held for a longer duration to acquire the information. To maintain attention in these cases, the warning needs to have qualities that generate interest, so that the person is willing to maintain attention to it. The effort necessary to acquire the information should be as little as possible. Thus, a goal is to enable the information to be grasped as easily as possible. Some of the same design features that facilitate the switch of attention also help to maintain attention. For example, large print not only attracts attention but also tends to increase legibility which makes the print easier to read.

Print legibility can be affected by numerous factors including choice of font, stroke width, letter compression and distance between them, resolution and justification (see Frascara, 2006). Although there is not much research to support an unequivocal preference for particular fonts, the general recommendation is to use relatively plain familiar alphanumerics. It is sometimes suggested that sans serif font like Helvetica, Futura and Univers for large text sizes and a serif font like Times, Times Roman and New Century Schoolbook be used for smaller-sized text. A chart with print sizes for expected reading distances in good and degraded conditions can be found in the ANSI (2002) Z535.4 warning standard.

Legibility is also benefited by high contrast between objects, such as text lettering relative to their background. Consider the poor contrast between the warning text stamped into the orange metal background of the pneumatic nail gun illustrated in Figure 12.6. In this instance, it is unlikely that workers will notice let alone maintain their attention with this particular warning. Black on white or the reverse (as seen on the container of an industrial cleaner depicted in Figure 12.7) has the highest contrast, but legibility can be adequate with other combinations such as black print on yellow (as in the ANSI Z535.4 CAUTION signal word panel) and white print on red (as in the ANSI



FIGURE 12.6 Embossed warning text in the metal of a nail gun.



FIGURE 12.7 Text on a container of industrial cleaner.

Z535.4 DANGER signal word panel). In such a situation, the custodial workers interacting with the cleaning materials shown in Figure 12.7 are more likely to see and maintain attention on the safety information than the construction workers interacting with the nail gun and its illegible warning illustrated in Figure 12.6.

People will more likely maintain attention if a warning is well designed (i.e. aesthetic) with respect to formatting and layout. Research suggests that people prefer warnings that are in a list



**FIGURE 12.8** Warning label on a riding lawnmower.

outline format as opposed to continuous prose text (Desaulniers, 1987). Also, text messages presented in all capital letters are worse than mixed-case text in glance legibility studies (Poulton, 1967), and centred-line formatting is worse than left-justified text (Hooper and Hannafin, 1986). Because individuals may decide it is too much effort to read large amounts of text, structured formatting could be beneficial in lessening the mental load and perception of difficulty. Formatting can make the visual display aesthetically pleasing to help hold people's attention on the material. Formatting can help process the information by *chunking* it into smaller units. Such text alterations should specifically benefit working memory because this memory system is assumed to be directed by the central executive, an attentional controller (Baddeley et al., 2009). Formatting can also show the structure or organisation of the material, thus making it easier to search for and assimilate the information into existing knowledge and memory housed in long-term memory systems (Hartley, 1994; Shaver and Wogalter, 2003). Figure 12.8 illustrates an example of a well-formatted warning housed on a riding lawnmower where black, left-justified, mixed-case text is presented in the form of a bullet list on a white background.

### Comprehension and Memory

Comprehension concerns understanding the meaning of something – in this case, the intended message of the warning. Comprehension may derive from several components: subjective understanding such as its hazard connotation, understanding of language and symbols, and an interplay with the individual's background knowledge. Background knowledge refers to relatively permanent long-term memory structure that may have resulted from exposure to safety information during

on-the-job training, during organised safety meetings or from reading operator manuals for tools used in the workplace. The following sections contain short reviews of some major conceptual research areas with respect to warnings and the comprehension stage.

### *Signal Words*

Aspects of a warning can convey a level of subjective hazard to the recipient. The ANSI (2002) Z535 standard recommends three signal words to denote different levels of hazard: DANGER, WARNING or CAUTION (see also Westinghouse Electric Corporation, 1981; FMC Corporation, 1985; Peckham, 2006). According to ANSI Z535, the DANGER panel should be used when serious injury or death *will* occur if the directive is not followed. A WARNING panel is used when serious injury or death *may* occur if the directive is not followed. The CAUTION panel is used when less severe personal injuries or property damage may occur if the directive is not followed. While the standard describes CAUTION and WARNING with different definitions, numerous empirical research studies indicate that people do not readily distinguish between the two. Although the term DEADLY has been shown in several research studies to connote significantly higher hazard than the standard's highest level DANGER, the use of DEADLY is not part of ANSI Z535 (see Wogalter and Silver, 1990, 1995; Wogalter et al., 1998a; Hellier and Edworthy, 2006).

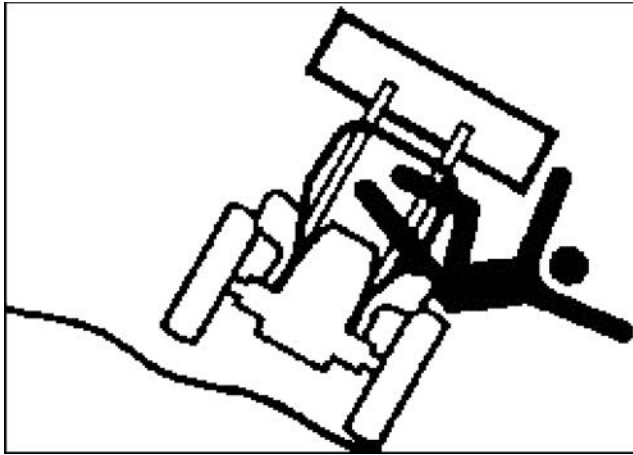
Consider the lawnmower warning label illustrated in Figure 12.8 as an example of an ANSI-style signal word panel. According to ANSI Z535, the signal word panels for DANGER, WARNING and CAUTION are assigned specific colours: red, orange and yellow, respectively. This assignment provides a form of redundancy due to the presence of more than one cue. However, most people do not reliably distinguish different levels of hazard associated with the colours orange and yellow (Chapanis, 1994; Wogalter et al., 1998a; Mayhorn et al., 2004c), and so the effect of redundancy is probably not very beneficial in this case. The signal word panels also contain the alert symbol (triangle/exclamation point), which indicates it is a warning (Wogalter et al., 1994a, 1998a).

### *Message Content*

The lawnmower warning illustrated in Figure 12.8 is an example of how a well-formatted industrial warning may fail due to inadequate warning content. Employees tasked with using a riding lawnmower must be made aware of *rollover* hazards where the operator might lose control when mowing on steep inclines. Rollover of the lawnmower onto the operator can result in potentially fatal crush injuries so it is imperative that operators understand the nature of the hazard from the warning. In general terms, the content of a warning message should include information about the hazard, instructions on how to avoid the hazard and the potential consequences if the hazard is not avoided (Wogalter et al., 1987). There are exceptions when the hazard is (1) general knowledge, (2) known from previous experience or (3) *open and obvious*, that is, apparent to everyone.

1. *Hazard information.* At a minimum, the warning should identify the safety problem. Oftentimes, however, warnings might require more information regarding the nature of the hazard and the mechanisms that produce it. In the case of Figure 12.9, the nature of the hazard is not described. Nowhere does the term *rollover* or *crush hazard* appear.
2. *Instructions.* Warnings should instruct people about what to do or not do. The instructions should be specific inasmuch as reasonable to tell what exactly should be done or avoided. The statement 'operate only on slopes you can back up and never on slopes greater than 15 degrees' does not tell that there is an injury potential due to rollover. In this case, without more information, operators are left making inferences which may be partly or wholly incorrect (Laughery et al., 1993a; Laughery and Paige-Smith, 2006).
3. *Consequences.* Consequences information concerns what could result. In Figure 12.9, *injury or death* is provided yet this statement is not sufficient to keep people from making incorrect inferences. As this warning illustrates, a common shortcoming in warning text





**FIGURE 12.9** Rollover symbol that might supplement the warning label on a riding lawnmower.

is that consequences-related information is not explicit, that is, lacking important specific details (Laughery et al., 1993a; Laughery and Paige-Smith, 2006). For instance, consider the consequences statement ‘Avoid serious injury or death’ from Figure 12.9 in the context of the resultant crush hazard due to rollover. This statement is insufficient by itself as it does not tell what kind of injury could occur. The operator might believe contact with the components of the motor might result in a thermal burn rather than thinking that it could be something more severe, like a crush resulting in loss of a limb and perhaps death. In a later section, the specification of severe consequences is discussed as a factor in motivating compliance behaviour.

### *Symbols*

Safety symbols such as the one depicted in Figure 12.9 may also be used to communicate the earlier-mentioned information regarding the rollover hazard in lieu of or in conjunction with text statements (e.g. Zwaga and Easterby, 1984; Young and Wogalter, 1990; Wolff and Wogalter, 1998; Dewar, 1999; Mayhorn et al., 2004b; Mayhorn and Goldsworthy, 2007, 2009). Not only might such a symbol act to tie the nature of the hazard to important instructions, but also the non-language attributes of symbols can contribute to understanding when illiterates or non-readers of the primary language are part of the target audience.

Comprehension is important for effective safety symbols (Dewar, 1999). Symbols that directly represent concepts are preferred because they are usually better comprehended than more abstract symbols (Wolff and Wogalter, 1993; Magurno et al., 1994; Wogalter et al., 2006). With abstract and arbitrary symbols (Sojourner and Wogalter, 1997, 1998; Wogalter et al., 1997; Lesch, 2003), the meaning typically has to be learned via training.

What is an acceptable level of comprehension for safety symbols? In general, symbols should be designed to have the highest level of comprehension attainable. The ANSI (2002) Z535 standard suggests a goal of at least 85% comprehension using a sample of 50 representative individuals from the target audience for a symbol to be used without accompanying text. If 85% cannot be achieved, the symbol may still have utility (e.g. for attention capture) as long as it is not likely to be misinterpreted. According to the ANSI (2002) Z535 standard, an acceptable symbol must produce less than 5% critical confusions (opposite meaning or a meaning that would produce unsafe behaviour). For example, a critical confusion might arise if lawnmower operators seeing the rollover symbol shown in Figure 12.9 misinterpret it to mean that sharp turns are acceptable as long as someone counter-balances the mower during operation. ISO (2001) has similar comprehension criteria (see Deppa, 2006; Peckham, 2006).



Repeated exposure to an unchanged warning over time will result in it being less effective not only in switching attention but also for maintaining attention (Kim and Wogalter, 2009). As mentioned earlier, even a well-designed warning will eventually become habituated if repeatedly encountered. Fortunately, habituation as a memory concept implies that the person has learned some amount of information from the warning to *know* to ignore it. Unfortunately, only part of the warning may actually be known. Some techniques for reducing habituation include (1) using salient features and (2) periodically varying the warning's appearance (and content, if feasible and appropriate).

Although individuals may have knowledge about a hazard, they may not be aware of it at the time they are at risk. Workers have vast stores of knowledge in long-term memory based on an accumulation of experience in their place of work. Despite this amazing memory storage space, at any given time, only a small portion of it is consciously available. As people are doing their work-related tasks, their minds are not always actively accessing risk information. Thus, while a person may have some or an extensive store of risk knowledge, this information and related knowledge may not be activated unless there is an external cue to activate it. Consider the potential threat to health-care workers from severe hypersensitivity to latex as might be encountered with latex gloves (Vredenburg et al., 2006). Because they are focused on patient care and repeatedly exposed to warning information on boxes of latex gloves, health-care workers may experience habituation to a warning label where it is infrequently noticed. But its presence is better than its absence, as, for example, it may serve as a reminder to some persons susceptible to latex hypersensitivity. So despite habituation, the presence of a warning may serve to cue relevant hazard information.

In summary, information in long-term memory can be cued by the presence of a warning and bring forth related, previously dormant knowledge into conscious awareness. Reminders may be appropriate in situations (1) where the hazard is infrequently encountered in which forgetting may be an issue and (2) when there are foreseeable distractions or high task-load involvement (e.g. patient care) that could pull attention away from normative hazard considerations.

### *Level of Knowledge*

The levels of knowledge and understanding of the warning recipients should be taken into consideration. Three cognitive characteristics of receivers are important: language skill, reading ability and technical knowledge.

With regard to the work environment, it is not unusual for workers to be given textual warnings beyond their reading skill. Consider the text from the cleaning product illustrated in Figure 12.8:

#### PRECAUTIONS – READ BEFORE USING

POISON: [skull & crossbones icon in red] MAY BE FATAL OR CAUSE PERMANENT DAMAGE IF SWALLOWED. CAUSES SEVERE BURNS TO EYES AND SKIN. Contains sodium hydroxide (caustic soda-lye). May cause blindness. Avoid contact with skin, eyes, mouth, and clothing.

ALWAYS KEEP OUT OF REACH OF CHILDREN. IMMEDIATELY GIVE FIRST AID. EYES: Rinse immediately with water. Remove contact lenses; then flush eyes with water for at least 20 minutes. IF SWALLOWED: Rapidly rinse mouth. Then immediately drink a glassful of milk or water. Do not induce vomiting. IF ON SKIN: Remove affected clothing. Flush with water for 15 minutes. THEN, IMMEDIATELY CALL POISON CENTER, PHYSICIAN OR EMERGENCY ROOM.

KEEP OUT OF REACH OF CHILDREN.

In general, reading levels should be as low as feasible. For the general population, the reading level probably should be approximately the skill level of grades 4–6 (expected ability of 10–12-year-old readers). In this instance, it is unlikely that the average worker will understand chemical references to sodium hydroxide and *caustic soda-lye*. When submitted to a readability analysis using the Flesch (1948) readability formula, the warning text given earlier was scored at a grade level of 6.3, which exceeds the skill level listed. Thus, the custodian tasked with using

this product during the course of their work-related duties may be exposed to safety hazards because he or she did not understand the warning.

There are large numbers of functionally illiterate persons, even in some of the most technologically advanced countries. For example, in the United States, there are estimates of over 30 million functionally illiterate adults, and like everyone else, these individuals will likely have to work (Kutner et al., 2007). Thus, successful warning communication may require more than simply keeping reading levels to a minimum. The use of symbols, speech warnings and special training programs may be beneficial adjuncts. Also, a related consideration is that different subgroups within a population may speak and read different languages. Because of increasing international trade and travel and the need to cross language barriers, this problem might require the use of multiple languages, graphics and transmission through multiple methods (Lim and Wogalter, 2003). An example of a multilingual warning is illustrated in Figure 12.1, which depicts safety information on a heat gun used to remove wall paper and paint. It shows a pictorial of a fire and text in both English and French, and further on the right side is Spanish.

Despite considerations at the minimal end, reading levels should be consistent with the reading abilities and level of knowledge associated with the previous training of the receivers. Such variables are often dependent on the type of occupation. A warning to trained health-care professionals such as nurses, physicians and pharmacists should use standard verbiage expected by that population. These technical experts have a more complete understanding of domain-specific hazards and can perform their jobs better with area-appropriate technical data. In contrast, a tyre salesperson or professional installer of tyres cannot be expected to have extensive training on the hazards and warnings associated with tyre construction and how this might impact proper installation. Training on the topic is likely to be no more than a short course or two, and likely less, such as on the job training. Here the warnings might not be much different in the level of difficulty than those transmitted to the public. With regard to warning systems, different components of the warning system can be used to communicate to different groups.

## **Beliefs and Attitudes**

*Beliefs and attitudes* is the next major stage of the C-HIP model. Beliefs refer to an individual's knowledge that is accepted as true (although some of it may not actually be true.) It is related to the previous stage in that beliefs are formed based on the examination of ideas stored in memory. In some respects, beliefs tend to be more global and overarching compared to specific memories. An attitude is similar to a belief except it includes more affect or emotional involvement.

A worker's benign experiences with a potentially hazardous tool can produce beliefs that a product is safer than it is. Consider the tree removal specialist who has operated the woodchipper without incident for years. A lackadaisical belief in equipment operation quickly changes after being involved in some way with (or seeing) a serious injury event. According to the C-HIP model, a warning will be successfully processed at the beliefs-and-attitudes stage if the message concurs (or at least is not discrepant) with the receiver's current beliefs and attitudes. However, if the warning information does not concur, then beliefs and attitudes may need to be altered before a person will be motivated to carry out the warning's directed behaviour. The message and/or other information need to be persuasive to override existing incorrect beliefs and attitudes. Methods of persuasion are commonly used in advertising and have been empirically explored in the social and cognitive psychology literatures. Recent applications of persuasion include the design and implementation of persuasive technology that includes the manipulation of computerised systems for the purpose of influencing attitudes and beliefs (Fogg, 2003). Sometimes not only unequivocal and explicit statements can be used to persuade but also the features of the warning may convey a higher-level importance. Such persuasion is important when a product is more hazardous than people believe. Discussed in the following paragraphs are several relevant and interrelated factors associated with the beliefs and attitudes stage: hazard perception, familiarity, prior experience and relevance (see DeJoy, 1999; Riley, 2006; Vredenburg and Zackowitz, 2006).

Hazard perceptions influence processing at the beliefs-and-attitudes stage. The greater the perceived hazard, the more responsive people will be to warnings, as in looking for, reading and complying with them. The converse is also true. People are less likely to look for, read or comply with a warning for products that they do not believe are hazardous. Perceived hazard is closely tied to beliefs about injury severity. People perceive a product is more hazardous and act more cautiously when injuries could be severe (Wogalter et al., 1999). Interestingly, however, injury likelihood is a much less important factor in perceptions of risk or hazard for consumer products (Wogalter et al., 1991, 1993). Thus, the operator of the woodchipper will probably be swayed to comply with safety information if the extreme, though thankfully infrequent, consequence of losing a limb is communicated effectively.

Familiarity beliefs are formed from past similar experiences stored in memory. It is the belief that most everything that needs to be known about a product or situation is already known. A person believing that they are adequately familiar with a product might assume that a different, but similar, product operates in the same way and has the same hazards (which may not be true), reducing the likelihood that he or she will look for or read a warning (Godfrey and Laughery, 1984; Goldhaber and deTurck, 1988; Wogalter et al., 1991). For example, an employee familiar with one model of woodchipper may assume that a new piece of equipment possesses the same hydraulic feed system for capturing debris and delivering it to the cutting wheel. If, in fact, the feed system of the new equipment is much faster than that of the older device, employees may be at risk for having their limbs pulled into the device while loading debris due to this unexpected hazard and the mistaken belief that they are fast enough to avoid getting *sucked into the machine*.

Research indicates that hazard perception is more important than familiarity with respect to warnings (Wogalter et al., 1991). This is probably due to two factors. First, people more familiar with a situation or product may have more knowledge about the hazards and how to avoid them. Second, greater use also tends to increase exposure to warnings, which increases the opportunity to be influenced by them.

Related to familiarity is prior experience. The concepts are somewhat different in that familiarity is a belief (that may or may not be true), and prior experience is an objective quantity that could potentially be measured. Prior experience can be influential in hazard perceptions. Having experienced some form of injury or having personal knowledge of someone else being injured enhances the hazard perceptions (Wogalter et al., 1993). For instance, product users who were personally familiar with the hazards associated with cleaning solutions or who were aware of injuries to someone else were able to produce more effective hazard avoidance strategies (Mayhorn et al., 2004a). Similarly, the lack of such experiences may lead to underestimating dangers, or not thinking about them at all. Warnings that give vivid explicit consequences may convince people to change beliefs when they have inappropriate low levels of perceived hazard.

Perceived relevance is the belief that something is applicable to the person. If the individual does not believe the warning is relevant to them, then the warning may fail to fulfil its intended purpose. The individual may instead attribute the warning as being directed to others and not to him- or herself. For example, a truck driver transporting containers of pesticides to an agricultural community may mistakenly assume that warnings on the container are meant solely for farm workers. In such a case, the truck driver may erroneously believe that his limited exposure to the pesticide during the loading and offloading process is safe and that hand washing is not necessary. One way to counter this is to personalise the warning so that it gets directed to relevant users and conveys facts that indicate that it is relevant (Wogalter et al., 1994b).

A point related to beliefs and attitudes and more specifically, familiarity, concerns the problem of experts overestimating what people know, which in turn may affect what kinds of warnings are produced (Laughery, 1993). Experts in a domain can be so facile with their knowledge about a topic that they fail to realise that non-experts do not have similar knowledge. What is *obvious* to them may not be as equally obvious to end users. Without operator or end user input into the design of warnings, there may be a tendency to produce warnings that fail to meet the needs of workers.

## Motivation

Motivation energises the individual to carry out an activity and serves as an essential component that links attitudes to actual behaviour as described in the theory of planned behaviour described by Ajzen (1991). Some of the main factors that can influence the motivation stage of the C-HIP model are cost of compliance, severity of injury, social influence and stress. These topics are discussed later in text.

Compliance generally requires that people take some action, and usually there are costs associated with doing so. The costs of complying may include time and effort to carry out the behaviour (Wogalter et al., 1987, 1989). When people perceive the costs of compliance to be too high, they are less likely to perform the safety behaviour. This problem is commonly encountered in warnings with instructions directing behaviours that are inconvenient, difficult, uncomfortable or occasionally impossible to carry out. For example, *long haul* truck drivers operate for extended periods of time that require them to use sleeper cabin berths inside their trucks (Darwent et al., 2012). To meet tight work deadlines, these drivers frequently operate in teams of two such that one can drive while his/her partner is sleeping. For safety purposes, a sleeper restraint system (e.g. seat belt for prone *off duty* driver in the berth) is provided. Unfortunately, death or injury due to ejection during a crash can occur. Anecdotally, many truck drivers have expressed an unwillingness to use the restraint system due to discomfort while sleeping. One way to reduce this cost of compliance is to make the directed behaviour easier and more comfortable to perform. Perhaps sleeper restraints that are designed using thicker padding might reduce discomfort and increase the likelihood of using this safety equipment. Such a finding would be consistent with previous research that indicates that the provision of protective devices such as gloves during tool operation reduces the costs of compliance (Wogalter et al., 1989; Dingus et al., 1991).

The costs of non-compliance can also exert a powerful influence on compliance motivation. With respect to warnings, the main cost for non-compliance is severe injury consequences. Previous research suggests that people report higher willingness to comply with warnings when they believe there is high probability for incurring a severe injury (e.g. Wogalter et al., 1991, 1993, 1999e). In this fashion, perhaps long haul truckers could be enticed to use their sleeper berth restraint systems if they are told it is similar to driving without a seat belt. Warning messages accompanied by explicit images depicting the fatal consequences of being ejected from a vehicle during a crash should be effective in further illustrating the costs of non-compliance.

Another motivator is social influence (Wogalter et al., 1989; Edworthy and Dale, 2000). When people see others comply with a warning, they are more likely to comply themselves. Likewise, seeing others not comply lessens the likelihood of compliance. Other factors affecting motivation are time stress (Wogalter et al., 1998b) and mental workload (Wogalter and Usher, 1999). Under high stress and workload, competing activities take resources away from processing warning information.

## BEHAVIOUR

The last stage of the sequential process is for individuals to carry out the warning-directed safe behaviour. Behaviour is one of the most important measures of warning effectiveness (Silver and Braun, 1999; Kalsher and Williams, 2006). Warnings do not always change behaviour because of processing failures at earlier stages. Most research in this area focuses on the factors that affect compliance likelihood, including those that enhance safety behaviour and those that do not.

Some researchers have used *intentions to comply* as the method of measurement because it is usually quite difficult to conduct behavioural tests. The difficulties include the following: (1) researchers cannot expose participants to real risks because of ethical and safety concerns; (2) events that could lead to injury are relatively rare; (3) the stimulus scenario must appear to have a believable risk, yet at the same time must be safe; and (4) running such research is costly in terms of time and effort. Nevertheless, compliance is an important criterion for determining which factors work better than others to boost warning effectiveness and, consequently, safe behaviour in the workplace.

Virtual reality or simulation may play a role in allowing research to be conducted in simulated conditions that avoid some of the problems provided earlier (Duarte et al., 2010). Also compliance can be measured indirectly. For example, determining whether protective gloves have been worn can be gleaned from whether they appear to be used or stretched in appearance (Wogalter and Dingus, 1999; Kalsher and Williams, 2006). Likewise, sleep restraint system use in truck drivers could be assessed at a global level by determining whether the belt demonstrates wear and tear from use (as opposed to retaining the neatly folded, plastic wrapped factory configuration). (See Chapter 8 for further discussion of the role of simulation tools in E/HF research.)

### **Receiver Variables**

The receiver's characteristics and task workload can affect warning effectiveness (Young et al., 1999). Indeed, evidence supporting this has already been discussed. Person variables (Rogers et al., 2000) such as the individuals' existing knowledge, beliefs and language skill were noted in earlier sections as affecting whether and how a warning is processed. Likewise, the demographic characteristics of workers should be considered. Previous research findings indicate that age-related declines in sensory and cognitive processing can affect warning processing particularly in attention switch and memory/comprehension stages (see Mayhorn and Podany, 2006; McLaughlin and Mayhorn, 2014). Because the world's workforce is ageing at an unprecedented rate, older employees may be placed at a heightened level of risk in certain occupations (Foster-Thompson and Mayhorn, 2012). For instance, older farmers have been shown to be differentially susceptible to tractor-related injuries due to decreased reaction time and other factors where age-related decrements have been observed (McLaughlin and Mayhorn, 2011). Thus, efforts to protect these particular workers would benefit from the development of age-appropriate warnings that take cognitive, motoric and perceptual abilities/limitations into account.

Given the increasing diversity of the workforce, other demographic factors also need to be considered. In some studies, gender differences have been noted (see, e.g., Laughery and Brelsford, 1991; Smith-Jackson, 2006a,b) with women being somewhat more likely to look for and read warnings (e.g. Godfrey et al., 1983; LaRue and Cohen, 1987; Young et al., 1989; Tam and Greenfield, 2010). Other research indicates that risk perception varies by ethnicity such that Latino farm workers reported higher risk perception associated with the use pesticides than Americans of European descent (Smith-Jackson et al., 2010).

Finally, warning processing occurs in the context of other potential processing given other stimuli in the environment and the individual's ongoing and ever-changing work behaviour. Whether and how a warning is processed can depend on mental workload (Wogalter and Usher, 1999), time stress (Wogalter et al., 1998b) and processing strategy (deTurck and Goldhaber, 1988). Consider one factory employee working under strict time constraints to meet a mandatory-specified quota versus an employee who is paid by the hour. The first individual operating under time pressure is probably not in an information-seeking mode and therefore is less likely to fully process a warning compared to the second individual who is not under those constraints. When such task loading can be anticipated (e.g. in emergency situations or when employees are working under time pressure), the warning system may have to be highly salient to attract attention.

## **A LOOK INTO THE FUTURE: TWENTY-FIRST-CENTURY INCLUSIVE WARNING SYSTEMS**

To prepare for a future characterised by growing international trade and increased daily interaction between geographically separated co-workers, workplace warning systems currently in place must be adapted. To accommodate the needs of an increasingly diverse workforce as described earlier, inclusive warnings must be developed (Mayhorn et al., 2014). For instance, different cultural subgroups within the workforce of a single multinational corporation may speak and read different primary languages. Thus, an effective workplace warning must be able to cross cultural and language barriers.



One such attempt within the United States was assessed by Lim and Wogalter (2003), who concluded that culturally inclusive warnings require the use of multiple languages, combined graphics and transmission through multiple methods to reach various subpopulations that receive it.

As previously addressed, one potential solution to overcoming the language barrier might include increased reliance on pictorial symbols in warnings. Despite the apparent potential benefits to using symbols to convey hazard information, there have been a number of studies that show cultural differences in how people interpret the meaning of symbols. One example of such cultural differences was documented when ANSI symbols were tested for comprehension in Ghana. Severe interpretation discrepancies were noted for a number of symbols and their intended meanings (Smith-Jackson and Essuman-Johnson, 2002). Other research found that drivers from Canada, Israel, Finland and Poland displayed large comprehension differences among traffic signs (Shinar et al., 2003). Likewise, residents of Hong Kong had difficulty interpreting the meaning of some industrial signs used in Mainland China (Chan and Ng, 2010). Thus, symbols should be tested for comprehension within the intended target audience (even when the perceived subcultures are geographically proximal to one another) prior to deployment in a workplace warning system.

Just as the content of future warning systems might change, avenues for delivery of safety information are likely to change as well. With technology becoming ubiquitous in the workplace, safety professionals can begin to use the attributes of technology to create more effective dynamic warnings (Wogalter and Mayhorn, 2005). Not only could such use of technology combat information processing hurdles (i.e. habituation) to warning compliance identified in the content of the C-HIP model described earlier, dissemination could be tailored to meet the known characteristics of workers within a specific location. For instance, cellular telephone usage is commonplace, and many of these devices are equipped with global positioning systems. Because users enter their personal information (e.g. names and language preference) into these devices and carry them on their persons, an avenue for customised warning dissemination is now open where safety information can be delivered close in time and proximity to when it is needed. Such tailoring might also be used to impact the beliefs and attitudes of workers via persuasive technology (Fogg, 2003). Likewise, other technology such as computers and tablets are also available throughout the modern workplace for dissemination of warning information and training materials. As technology in the workplace continues to evolve, it is likely that warning systems will evolve to be more effective in promoting safe behaviour. Ultimately, the success of these future warning systems is dependent on how well safety professionals exploit these technological opportunities.

## SUMMARY AND UTILITY OF THE C-HIP MODEL

The review of the warning literature as applied to the workplace environment was organised around the C-HIP model. This model is valuable in describing the information processing steps that occur when safety information is encountered and organising a large amount of previous research on the topic. Furthermore, it can also be a valuable tool in systematising the assessment process to help determine why a warning is not effective. It can aid in pinpointing where the bottlenecks in processing may be occurring and suggest solutions to allow processing to continue to subsequent stages. Moreover, workplace warning effectiveness testing can be performed using methods similar to those used in research settings. Evaluations of the processing can be directed to any of the stages described in the C-HIP model: source, channel, environment, delivery, attention, comprehension, attitudes and beliefs, motivation, behaviour and receiver variables. Some of the methods for doing this evaluation are briefly described here.

Evaluating the source necessitates an attempt to determine whether responsible parties such as employers and equipment manufacturers have documented the potential hazards and issued warnings. It is fundamental that workplace equipment manufacturers should analyse their products to determine whether there are foreseeable potential hazards associated with their use and misuse. When hazards are discovered, manufacturers have an obligation to employ methods to try to control



the hazards to reduce personal injury and property damage. If a manufacturer is going to sell a product in which the hazard has not been eliminated through design or physical guarding, then it should provide effective warning(s) to end users such as employers forced to use these devices to earn a living.

Efforts to evaluate the channel of warning delivery mainly assess how safety information is sent to end users. What media and modalities are being used and are those adequate? Similarly, assessment regarding delivery asks whether end users receive the warnings. If not, other channels of distribution of warning materials may need to be considered. As mentioned earlier, the concept of *cascading responsibility* in the chain of commerce requires that the equipment manufacturers, intermediaries (e.g. distributors and retailers) and employers share a responsibility to provide workers with needed safety information (Williams et al., 2006). Thus, warning delivery must be assessed at each level of responsibility.

To assess attention switch, the main question is whether workers see or hear the warnings. The answer could involve placing a warning in the workplace and having people carry out a relevant task then asking them later whether they saw it (McGrath, 2011). Eye movement and response time paradigms represent other methods used to measure what people tend to look at and how quickly.

To assess comprehension, there are several well-established methodologies involving memory tests, open-ended response tests, structured interviews, etc. These assessment instruments can be valuable for determining what information was or was not understood and for suggesting revisions to warning text or symbols. To assess beliefs and attitudes, a questionnaire could be used to determine people's pre-existing beliefs on the topics of perceived hazard and familiarity with the tool, task or environment. For example, if workers' perceived hazard is too low for a situation, greater persuasiveness may be needed to promote warning compliance.

To assess motivation, measures of behavioural intentions can be used. Low intentions to comply may indicate that consequence information should be enhanced (e.g. by being more explicit) or that cost of compliance should be reduced. To assess behavioural compliance, systematic observation can be used in both lab and field settings. As mentioned earlier, measurement of behavioural compliance is generally more difficult than any of the other methods. It may involve ethical issues such as participants' exposure to risk. However, in situations where the negative consequences are substantial, the effort and resources may be warranted. Sometimes behavioural intentions are measured as a proxy for overt behavioural compliance – but, some caution should be exercised, as noted earlier.

In summary, workplace warning interaction can be conceptualised using the C-HIP model and investigated using the methods (and others) described earlier in a systematic manner. By determining the specific causes of a warning's failure, potential solutions can be generated and assessed. Resources can then be directed at fixing any shortcomings that limit observed warning effectiveness. In this manner, employers and employees can benefit from a safer, more productive workplace environment.

## REFERENCES

- Ajzen, I. 1991. Theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, 50(2), 179–211.
- ANSI. 2002. Accredited standards committee on safety signs and colors. Z535.1-5. Arlington, VA: National Electrical Manufacturers Association.
- Baddeley, A., Eysenck, M. W. and Anderson, M. C. 2009. *Memory*. New York: Psychology Press.
- Baldwin, C., Spence, C., Bliss, J., Brill, C. J., Wogalter, M. S., Mayhorn, C. B. and Ferris, T. 2012. Multimodal cueing: The relative benefits of the auditory, visual, and tactile channels in complex environments. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Barlow, T. and Wogalter, M. S. 1993. Alcoholic beverage warnings in magazine and television advertisements. *Journal of Consumer Research*, 20, 147–155.

- Bzostek, J. A. and Wogalter, M. S. 1999. Measuring visual search time for a product warning label as a function of icon, color, column, and vertical placement. *Proceedings of the Human Factors and Ergonomics Society*, 43, 888–892.
- Chan, A. H. S. and Ng, A. W. Y. 2010. Investigation of guessability of industrial safety signs: Effects of prospective-user factors and cognitive sign features. *International Journal of Industrial Ergonomics*, 40(6), 689–697.
- Chapanis, A. 1994. Hazards associated with three signal words and four colours on warning signs. *Ergonomics*, 37, 265–275.
- Clark, J. M. and Paivio, A. 1991. Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210.
- Cohen, H. H., Cohen, J., Mendat, C. C. and Wogalter, M. S. 2006. Warning channel: Modality and media. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 123–134. Mahwah, NJ: Lawrence Erlbaum Associates.
- Conzola, V. C. and Wogalter, M. S. 2001. A communication-human information processing (C-HIP) approach to warning effectiveness in the workplace. *Journal of Risk Research*, 4(4), 309–322.
- Cox, E. P. III and Wogalter, M. S. 2006. Warning source. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 111–122. Mahwah, NJ: Lawrence Erlbaum Associates.
- Darwent, D., Roach, G. and Dawson, D. 2012. How well do truck drivers sleep in cabin berths? *Applied Ergonomics*, 43(20), 442–446.
- DeJoy, D. M. 1999. Beliefs and attitudes. In M. S. Wogalter, D. M. DeJoy and K. R. Laughery (eds.), *Warnings and Risk Communication*, pp. 183–219. London, U.K.: Taylor & Francis Group.
- Deppa, S. W. 2006. U.S. and international standards for safety symbols. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 477–486. Mahwah, NJ: Lawrence Erlbaum Associates.
- Desaulniers, D. R. 1987. Layout, organization, and the effectiveness of consumer product warnings. *Proceedings of the Human Factors Society*, 31, 56–60.
- deTurk, M. A., Chih, I. H. and Hsu, Y. R. 1999. Three studies testing the effects of role models on product users' safety behaviour. *Human Factors*, 41(3), 397–412.
- deTurk, M. A. and Goldhaber, G. M. 1988. Consumers' information processing objects and effects of product warning. *Proceedings of the Human Factors Society*, 32, 445–449.
- Dewar, R. 1999. Design and evaluation of graphic symbols. In H. J. G. Zwaga, T. Boersema and H. C. M. Hoonhout (eds.), *Visual Information for Everyday Use: Design and Research Perspectives*, pp. 285–303. London, U.K.: Taylor & Francis Group.
- Dingus, T. A., Hathaway, J. A. and Hunn, B. P. 1991. A most critical warning variable: Two demonstrations of the powerful effects of cost on warning compliance. *Proceedings of the Human Factors Society*, 35, 1034–1038.
- Duarte, M. E. C., Rebelo, F. and Wogalter, M. S. 2010. Virtual reality and its potential for evaluating warning compliance. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 20(6), 526–537.
- Edworthy, J. and Dale, S. 2000. Extending knowledge of the effects of social influence in warning compliance. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society*, Vol. 4, pp. 107–110. Santa Monica, CA: Human Factors and Ergonomics Society.
- Flesch, R. 1948. A new readability yardstick. *Journal of Applied Psychology*, 32, 2211.
- FMC Corporation. 1985. *Product Safety Sign and Label System*. Santa Clara, CA: FMC Corporation.
- Frantz, J. P. and Rhoades, T. P. 1993. A task analytic approach to the temporal placement of product warnings. *Human Factors*, 35, 719–730.
- Frantz, J. P., Rhoades, T. P. and Lehto, M. R. 1999. Practical considerations regarding the design and evaluation of product warnings. In M. S. Wogalter, D. M. DeJoy and K. R. Laughery (eds.), *Warnings and Risk Communication*, pp. 291–311. London, U.K.: Taylor & Francis Group.
- Frascara, J. 2006. Typography and the visual design of warnings. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 385–406. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fogg, B. J. 2003. *Persuasive Technology*. San Diego, CA: Morgan Kaufmann.
- Foster-Thompson, L. and Mayhorn, C. B. 2012. Aging and technology. In J. W. Hedge and W. Borman (eds.), *Oxford Handbook of Work and Aging*, pp. 341–364. New York: Oxford University Press, Inc.
- Godfrey, S. S., Allender, L., Laughery, K. R. and Smith, V. L. 1983. Warning messages: Will the consumer bother to look? *Proceedings of the Human Factors Society*, 27, 950–954.
- Godfrey, S. S. and Laughery, K. R. 1984. The biasing effect of familiarity on consumer's awareness of hazard. *Proceedings of the Human Factors Society*, 28, 483–486.

- Goldhaber, G. M. and deTurck, M. A. 1988. Effects of consumer's familiarity with a product on attention and compliance with warnings. *Journal of Products Liability*, 11, 29–37.
- Guchelaar, H., Colen, H. B., Kalmeyer, M. D., Hudson, P. T. and Teepe-Twiss, I. M. 2005. Medication errors: Hospital pharmacist perspective. *Drugs*, 65(13), 1735–1746.
- Hartley, J. 1994. *Designing Instructional Text*, 3rd edn. London, U.K.: Kogan Page/East Brunswick, NJ: Nichols.
- Hatem, A. and Letho, M. 1995. Effectiveness of glue odor as a warning signal. *Ergonomics*, 38(11), 2250–2261.
- Helander, M. 1997. The human factors profession. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics*, 2nd edn., pp. 3–16. New York: Wiley.
- Hellier, E. and Edworthy, J. 2006. Signal words. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 407–417. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hooper, S. and Hannafin, M. J. 1986. Variables affecting the legibility of computer generated text. *Journal of Instructional Development*, 9, 22–28.
- ISO. 2001. Graphical symbols – test methods for judged comprehensibility and for comprehension, ISO 9186. Geneva, Switzerland: International Organization for Standards.
- Kalsher, M. J. and Williams, K. J. 2006. Behavioural compliance: Theory, methodology, and results. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 289–300. Mahwah, NJ: Lawrence Erlbaum Associates.
- Keon, T. L. and McDonald, B. 1982. Job satisfaction and life satisfaction: An empirical evaluation of their inter-relationship. *Human Relations*, 35, 167–180.
- Kim, S. and Wogalter, M. S. 2009. Habituation, dishabituation, and recovery effects in visual warnings. *Proceedings of the Human Factors and Ergonomics Society*, 53, 1612–1616.
- Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y. and Dunleavy, E. 2007. Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy (NCES 2007-480). Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- LaRue, C. and Cohen, H. 1987. Factors influencing consumer's perceptions of warning: An examination of the differences between male and female consumers. *Proceedings of the Human Factors Society*, 31, 610–614.
- Laughery, K. R. and Hammond, A. (1999). Overview. In M. S. Wogalter, D. M. DeJoy and K. R. Laughery (eds.), *Warnings and Risk Communication*, pp. 3–13. Philadelphia, PA: Taylor & Francis Group.
- Laughery, K.R. 8–13 July 1993. Everybody knows: Or do they? *Ergonomics in Design*. pp. 8–13.
- Laughery, K. R. 2006. Safety communications: Warnings. *Applied Ergonomics*, 37, 467–478.
- Laughery, K. R. and Brelsford, J. W. 1991. Receiver characteristics in safety communications. *Proceedings of the Human Factors Society*, 35, 1068–1072.
- Laughery, K. R. and Paige-Smith, D. 2006. Explicit information in warnings. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 419–428. Mahwah, NJ: Lawrence Erlbaum Associates.
- Laughery, K. R., Vaubel, K. P., Young, S. L., Brelsford, J. W. and Rowe, A. L. 1993. Explicitness of consequence information in warning. *Safety Science*, 16, 597–613.
- Laughery, K. R. and Wogalter, M. S. 1997. Risk perception and warnings. In G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics*, 2nd edn. New York: Wiley-Interscience.
- Laughery, K. R. and Wogalter, M. S. 2006. Designing effective warnings. In R. Williges (ed.), *Reviews of Human Factors and Ergonomics*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Laughery, K. R., Young, S. L., Vaubel, K. P. and Brelsford, J. W. 1993. The noticeability of warnings on alcoholic beverage containers. *Journal of Public Policy and Marketing*, 12, 38–56.
- Lehto, M. R. and Miller, J. M. 1986. *Warnings: Fundamentals, Design and Evaluation Methodologies*, Vol. 1. Ann Arbor, MI: Fuller Technical Publications.
- Lehto, M. R. and Salvendy, G. 1995. Warnings: A supplement not a substitute for other approaches to safety. *Ergonomics*, 38, 2155–2163.
- Lenorovitz, D. R., Karnes, E. W. and Lenoard, S. D. 2014. Mitigating certain hazards with just warnings: When and why such warnings are likely to fail. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 24(3), 275–297.
- Leonard, S. D., Otani, H. and Wogalter, M. S. 1999. Comprehension and memory. In M. S. Wogalter, D. M. DeJoy and K. R. Laughery (eds.), *Warnings and Risk Communication*, pp. 149–188. London, U.K.: Taylor & Francis Group.
- Lesch, M. F. 2003. Comprehension and memory for warning symbols: Age-related differences and impact of training. *Journal of Safety Research*, 34, 495–505.
- Lim, R. W. and Wogalter, M. S. 2003. Beliefs about bilingual labels on consumer products. *Proceedings of the Human Factors and Ergonomics Society*, 47, 839–843.

- Luczak, H., Hinrichsen, S. and Mutze-Niewohner, S. 2006. Humans in work system environment. In W. S. Marras and W. Karwowski (eds.), *The Occupational Ergonomics Handbook*, 2nd edn., pp. 5.1–5.52. London, U.K.: Taylor & Francis Group.
- Magurno, A., Wogalter, M. S., Kohake, J. and Wolff, J. S. 1994. Iterative test and development of pharmaceutical pictorials. *Proceedings of the 12th Triennial Congress of the International Ergonomics Association*, Vol. 4, pp. 360–362. Toronto, Canada.
- Marquardt, N., Gades, R. and Robelski, S. 2012. Implicit social cognition and safety culture. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22(3), 213–234.
- Mayhorn, C. B. and Goldsworthy, R. C. 2007. Refining teratogen warning symbols for diverse populations. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 79(6), 494–506.
- Mayhorn, C. B. and Goldsworthy, R. C. 2009. “New and improved”: The role text augmentation and the application of responses interpretation standards (coding schemes) in a final iteration of birth defects warnings development. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 85(10), 864–871.
- Mayhorn, C. B., Nichols, T. A., Rogers, W. A. and Fisk, A. D. 2004a. Hazards in the home: Using older adults’ perceptions to inform warning design. *Journal of Injury Control and Safety Promotion*, 11(4), 211–218.
- Mayhorn, C. B. and Podany, K. I. 2006. Warnings and aging: Describing the receiver characteristics of older adults. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 355–362. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mayhorn, C. B., Wogalter, M. S. and Bell, J. L. 2004b. Are we ready? Misunderstanding homeland security safety symbols. *Ergonomics in Design*, 12(4), 6–14.
- Mayhorn, C. B., Wogalter, M. S., Goldsworthy, R. C. and McDougald, B. R. 2014. Creating inclusive warnings: The role of culture in the design and evaluation of risk communications. In T. Smith-Jackson (ed.), *Cultural Ergonomics: Theories, Methods, and Applications*, pp. 97–128. Clermont, FL: Taylor & Francis Group.
- Mayhorn, C. B., Wogalter, M. S. and Shaver, E. F. 2004c. What does code red mean? *Ergonomics in Design*, 2(4), 12.
- Mazis, M. B. and Morris, L. A. 1999. Channel. In M. S. Wogalter, D. M. DeJoy and K. R. Laughery (eds.), *Warnings and Risk Communication*, pp. 99–121. London, U.K.: Taylor & Francis Group.
- McGrath, J. M. (2011). The role of equipment warning labels in the industrial workplace. *International Journal of Occupational Safety and Ergonomics*, 17(1), 49–60.
- McLaughlin, A. C. and Mayhorn, C. B. 2011. Avoiding harm on the farm: Human factors. *Gerontechnology*, 10(1), 26–37.
- McLaughlin, A. C. and Mayhorn, C. B. 2014. Designing effective risk communications for older adults. *Safety Science*, 61, 59–65.
- Mehlenbacher, B., Wogalter, M. S. and Laughery, K. R. 2002. On the reading of product owner’s manuals: Perceptions and product complexity. *Proceedings of the Human Factors and Ergonomics Society*, 46, 730–734.
- Olson, R., Grosshuesch, A., Schmidt, S., Gray, M. and Wipfli, B. 2009. Observational learning and workplace safety: The effects of viewing the collective behaviour of multiple social models on the use of personal protective equipment. *Journal of Safety Research*, 40(5), 383–387.
- Parker, D., Lawrie, M. and Hudson, P. 2006. A framework for understanding the development of organizational safety culture. *Safety Science*, 44, 551–562.
- Pavio, A. 1971. *Imagery and Verbal Processes*. New York: Holt, Rinehart and Winston.
- Peckham, G. M. 2006. ISO design standards for safety signs and labels. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 455–462. Mahwah, NJ: Lawrence Erlbaum Associates.
- Poulton, E. 1967. Searching for newspaper headlines printed in capitals or lower-case letters. *Journal of Applied Psychology*, 51, 417–425.
- Riley, D. M. 2006. Beliefs, attitudes, and motivation. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 289–300. Mahwah, NJ: Lawrence Erlbaum Associates.
- Robinson, G. S. and Casali, J. G. 1995. Auditory reverse alarms under hearing protectors for normal and hearing-impaired listeners. *Ergonomics*, 38(11), 2281–2299.
- Rogers, W. A., Lamson, N. and Rousseau, G. K. 2000. Warning research: An integrative perspective. *Human Factors*, 42, 102–139.
- Russ, A. L., Zillich, A. J., McManus, M. S., Doebbeling, B. N. and Saleem, J. J. 2012. Prescribers’ interactions with medication alerts at the point of prescribing: A multi-method, in situ investigation of the human-computer interaction. *International Journal of Medical Informatics*, 81(4), 232–243.
- Sanders, M. S. and McCormick, E. J. 1993. *Human Factors in Engineering and Design*, 7th edn. New York: McGraw-Hill.



- Shaver, E. F. and Wogalter, M. S. (2003). A comparison of older v. newer over-the-counter (OTC) nonprescription drug labels on search time accuracy. *Proceedings of the Human Factors and Ergonomics Society*, 47, 826–830.
- Shinar, D., Dewar, R. E., Summala, H. and Zakowski, L. 2003. Traffic symbol comprehension: A cross-cultural study. *Ergonomics*, 46(15), 1549–1565.
- Silver, N. C. and Braun, C. C. 1999. Behaviour. In M. S. Wogalter, D. M. DeJoy and K. R. Laughery (eds.), *Warnings and Risk Communication*, pp. 245–262. London, U.K.: Taylor & Francis Group.
- Smith-Jackson, T.L. 2006a. Culture and warnings. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 363–372. Mahwah, NJ: Lawrence Erlbaum Associates.
- Smith-Jackson, T. L. 2006b. Receiver characteristics. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 335–344. Mahwah, NJ: Lawrence Erlbaum Associates.
- Smith-Jackson, T. L. and Essuman-Johnson, A. 2002. Cultural ergonomics in Ghana, West Africa: A descriptive study of industry and trade workers' interpretations of safety symbols. *International Journal of Occupational Safety and Ergonomics*, 8(1), 37–50.
- Smith-Jackson, T.L., Wogalter, M. S. and Quintela, Y. 2010. Safety climate and risk communication disparities for pesticide safety in crop production by ethnic group. *Human Factors and Ergonomics in Manufacturing*, 20(6), 511–525.
- Sojourner, R. J. and Wogalter, M. S. 1997. The influence of pictorials on evaluations of prescription medication instructions. *Drug Information Journal*, 31, 963–972.
- Sojourner, R. J. and Wogalter, M. S. 1998. The influence of pictorials on the comprehension and recall of pharmaceutical safety and warning information. *International Journal of Cognitive Ergonomics*, 2, 93–106.
- Tam, T. and Greenfield, T. 2010. Do alcohol warning labels influence men's and women's attempts to deter others from driving while intoxicated? *Human Factors and Ergonomics in Manufacturing*, 20(6), 538–546.
- Thorley, P., Hellier, E. and Edworthy, J. 2001. Habituation effects in visual warnings. In M. A. Hanson (ed.), *Contemporary Ergonomics 2001*, pp. 223–228. London, U.K.: Taylor & Francis Group.
- U.S. Bureau of Labor Statistics. 2012. International comparisons of annual labor force statistics: Adjusted to U.S. concepts, 16 countries, 1970–2011. Washington, DC: U.S. Bureau of Labor Statistics.
- Vredenburg, A. G., Longden, S., Williams, K. J. and Kalsher, M. J. 2006. Evaluating latex glove container warnings in a realistic setting. *International Journal of Industrial Ergonomics*, 35(6), 559–568.
- Vredenburg, A. G. and Zackowitz, I. B. 2006. Expectations. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 345–354. Mahwah, NJ: Lawrence Erlbaum Associates.
- Westinghouse Electric Corporation. 1981. *Product Safety Label Handbook*. Trafford, PA: Westinghouse Printing Division.
- Williams, K. J., Kalsher, M. J. and Laughery, K. R. 2006. Allocation of responsibility for injuries. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 617–628. Mahwah, NJ: Lawrence Erlbaum Associates.
- Williamson, R. B. 2006. Fire warnings. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 701–710. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wogalter, M. S. 2006b. Communication-human information processing (C-HIP) model (Chapter 5). In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 51–61. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wogalter, M. S., Allison, S. T. and McKenna, N. 1989. Effects of cost and social influence on warning compliance. *Human Factors*, 31, 133–140.
- Wogalter, M. S., Barlow, T. and Murphy, S. 1995. Compliance to owner's manual warnings: Influence of familiarity and the task-relevant placement of a supplemental directive. *Ergonomics*, 38, 1081–1091.
- Wogalter, M. S., Brelsford, J. W., Desaulniers, D. R. and Laughery, K. R. 1991. Consumer product warnings: The role of hazard perception. *Journal of Safety Research*, 22, 71–82.
- Wogalter, M. S., Brems, D. J. and Martin, E. G. 1993. Risk perception of common consumer products: Judgments of accident frequency and precautionary intent. *Journal of Safety Research*, 24, 97–106.
- Wogalter, M. S. and Dingus, T. A. 1999. Methodological techniques for evaluating behavioural intentions and compliance. In M. S. Wogalter, D. M. DeJoy and K. R. Laughery (eds.), *Warnings and Risk Communication*, pp. 53–82. London, U.K.: Taylor & Francis Group.
- Wogalter, M. S., Godfrey, S. S., Fontenelle, G. A., Desaulniers, D. R., Rothstein, P. R. and Laughery, K. R. 1987. Effectiveness of warnings. *Human Factors*, 29, 599–612.
- Wogalter, M. S., Jarrard, S. W. and Simpson, S. W. 1994a. Influence of signal words on perceived level of product hazard. *Human Factors*, 36, 547–556.
- Wogalter, M. S., Kalsher, M. J., Frederick, L. J., Magurno, A. B. and Brewster, B. M. 1998a. Hazard level perceptions of warning components and configurations. *International Journal of Cognitive Ergonomics*, 2, 123–143.

- Wogalter, M. S., Kalsher, M. J. and Rashid, R. 1999d. Effect of signal word and source attribution on judgments of warning credibility and compliance likelihood. *International Journal of Industrial Ergonomics*, 24, 185–192.
- Wogalter, M. S. and Leonard, S. D. 1999. Attention capture and maintenance. In M. S. Wogalter, D. M. DeJoy and K. R. Laughery (eds.), *Warnings and Risk Communication*, pp. 123–148. London, U.K.: Taylor & Francis Group.
- Wogalter, M. S., Magurno, A. B., Dietrich, D. and Scott, K. 1999e. Enhancing information acquisition for over-the-counter medications by making better use of container surface space. *Experimental Aging Research*, 25, 27–48.
- Wogalter, M. S., Magurno, A. B., Rashid, R. and Klein, K. W. 1998b. The influence of time stress and location on behavioural compliance. *Safety Science*, 29, 143–158.
- Wogalter, M. S. and Mayhorn, C. B. 2005. Providing cognitive support with technology-based warning systems. *Ergonomics*, 48, 522–533.
- Wogalter, M. S., Racicot, B. M., Kalsher, M. J. and Simpson, S. N. 1994b. The role of perceived relevance in behavioural compliance in personalized warning signs. *International Journal of Industrial Ergonomics*, 14, 233–242.
- Wogalter, M. S. and Silver, N. C. 1990. Arousal strength of signal words. *Forensic Reports*, 3, 407–420.
- Wogalter, M. S. and Silver, N. C. 1995. Warning signal words: Connoted strength and understandability by children, elders, and non-native English speakers. *Ergonomics*, 38, 2188–2206.
- Wogalter, M. S., Silver, N. C., Leonard, S. D. and Zaikina, H. 2006. Warning symbols. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 159–176. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wogalter, M. S., Sojourner, R. J. and Brelsford, J. W. 1997. Comprehension and retention of safety pictorials. *Ergonomics*, 40, 531–542.
- Wogalter, M. S. and Usher, M. 1999. Effects of Concurrent Cognitive Task Loading on Warning Compliance Behavior. *Proceedings of the Human Factors and Ergonomics Society*, Vol. 43, pp. 106–110.
- Wogalter, M. S. and Vigilante, W. J. Jr. 2006. Attention switch and maintenance. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 245–266. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wogalter, M. S., Vigilante, W. J. and Baneth, R. C. 1998c. Availability of operator manuals for used consumer products. *Applied Ergonomics*, 29, 193–200.
- Wogalter, M. S. and Young, S. L. 1991. Behavioural compliance to voice and print warnings. *Ergonomics*, 34, 79–89.
- Wolff, J. S. and Wogalter, M. S. 1993. Test and development of pharmaceutical pictorials. *Proceedings of Interface*, 93(8), 187–192.
- Wolff, J. S. and Wogalter, M. S. 1998. Comprehension of pictorial symbols: Effects of context and test method. *Human Factors*, 40, 173–186.
- Young, S. L., Frantz, J. P. and Rhoades, T. P. 2006. Revisions of labeling for personal watercraft: Label development and evaluation. In M. S. Wogalter (ed.), *Handbook of Warnings*, pp. 723–738. Mahwah, NJ: Lawrence Erlbaum Associates.
- Young, S. L., Laughery, K. R., Wogalter, M. S. and Lovvoll, D. 1999. Receiver characteristics in safety communications. In W. Karwowski and W. S. Marras (eds.), *The Occupational Ergonomics Handbook*, pp. 693–706. Boca Raton, FL: CRC Press.
- Young, S. L., Martin, E. G. and Wogalter, M. S. 1989. Gender differences in consumer product hazard perceptions. *Proceedings of Interface 89*, pp. 73–78. Santa Monica, CA: Human Factors and Ergonomics Society.
- Young, S. L. and Wogalter, M. S. 1990. Comprehension and memory of instruction manual warnings: Conspicuous print and pictorial icons. *Human Factors*, 32, 637–649.
- Zwaga, H. J. G. and Easterby, R. S. 1984. Developing effective symbols or public information. In R. S. Easterby and H. J. G. Zwaga (eds.), *Information Design: The Design and Evaluation of Signs and Printed Material*. New York: John Wiley & Sons.



---

# 13 Evaluating Human-Computer Interaction

*Chris Baber*

## CONTENTS

|                                                               |     |
|---------------------------------------------------------------|-----|
| Introduction.....                                             | 360 |
| When and Where to ‘Evaluate’ .....                            | 360 |
| Defining ‘Usability’ and ‘User Experience’ .....              | 363 |
| What Does HCI Evaluation Involve?.....                        | 365 |
| Evaluation Implies Critique .....                             | 365 |
| Evaluation Involves Measurement .....                         | 365 |
| Evaluation Requires Comparison with a Referent Model.....     | 366 |
| Evaluation against Other Products .....                       | 367 |
| Evaluation against Design Targets.....                        | 367 |
| Evaluation against User and Organisational Requirements ..... | 368 |
| Evaluation against Standards.....                             | 368 |
| An Overview of Usability Evaluation Methods.....              | 369 |
| Analytic Methods .....                                        | 369 |
| Usability Inspection Methods .....                            | 370 |
| User Reports.....                                             | 371 |
| HCI in the Wild .....                                         | 372 |
| Participative Evaluation.....                                 | 373 |
| Overview of User Experience Methods .....                     | 373 |
| Reporting Evaluation .....                                    | 375 |
| Evaluating Future HCI.....                                    | 377 |
| Discussion.....                                               | 378 |
| References.....                                               | 379 |

Ergonomics/Human Factors has always played a key role in evaluating technology. Whilst the disciplines of human-computer interaction (HCI) and E/HF are distinct, in the same way as there is much in HCI that has evolved from E/HF, there is a need for E/HF practitioners to be versed in typical HCI approaches. Furthermore, there is a need to strike a balance between approaches to evaluation which might be pragmatically expedient (e.g. in terms of generating results over short time periods or with very few respondents) and those approaches which are rigorous (e.g. in terms of generating results in a controlled and testable manner). In this chapter, two approaches to the evaluation of HCI are considered. The usability approach focuses on the performance of goal-based tasks using computer interfaces and, consequently, addresses issues of productivity, for example in terms of metrics for efficiency and effectiveness in achieving a particular goal, and the level of satisfaction that a user might express concerning the computer interface. In contrast, the user experience approach focuses on factors such as pleasure and aesthetics and addresses computer gaming or

personal digital devices. Each approach has international standards associated with it, both in terms of ways to conduct evaluation and in terms of measures which can be applied. The chapter includes examples of methods associated with each type of evaluation.

## INTRODUCTION

Design, in HCI, is often a combination of artistic, software, engineering and human-centred approaches. Each approach has its own perspective on what it means to ‘design’ something and, more importantly, what it means to ‘evaluate’ these designs, but the basic assumption for this chapter is that design *must* involve evaluation. The purpose of evaluation is to ‘assess our designs and test our systems to ensure that they actually behave as we expect and meet the requirements of the users’ (Dix et al., 1993). Evaluation could range from the reaction of the audience (or user) to a product to the quantification of the reliability of the product’s performance, to the measurement of the performance of the person using the product, but all share a common challenge: how can we determine whether the design of a computer interface or digital product is acceptable to the people who will use it? Consider the examples in Figure 13.1. On the left is a handheld augmented reality viewer (and what is shown on its display) and on the right is a webpage. Would the same form of evaluation be appropriate to determine the acceptability of each of these applications? Obviously, we could simply ask users of each application, but this raises problems of subjective evaluation that have been explored in Chapter 4. For HCI, the problems of subjective evaluation stem from *which* questions you ask *which* users performing *which* task in *which* environment.

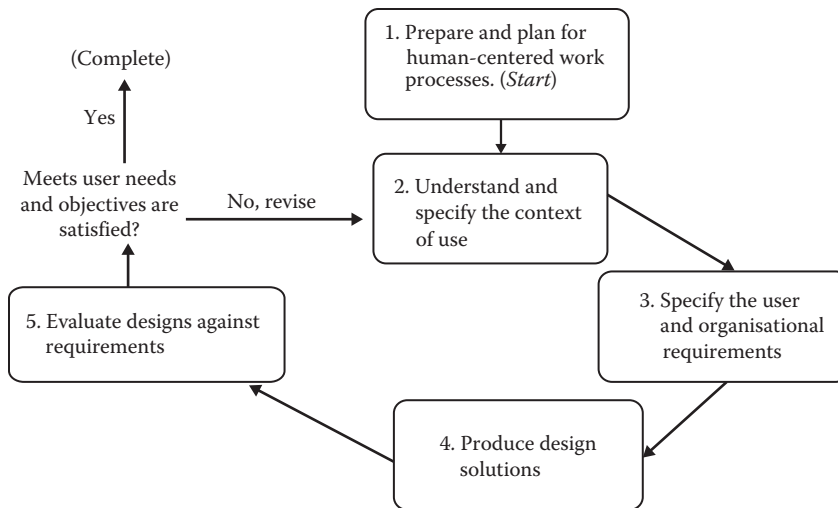
## WHEN AND WHERE TO ‘EVALUATE’

Before developing this discussion further, it is worth asking when one ought to conduct evaluation. The answer to this question rather depends on what one is measuring and where one is in the design life cycle. To start with the second point, the majority of books on software engineering and user interface design introduce the notion of a design life cycle. This is presented as a series of landmarks which a project will pass during its developmental life. In this chapter, we follow the lead of ISO 13407 (1999)\* (Figure 13.2). This iterative life cycle moves through the design to phase 5, which is the point at which the designer (or design team) considers whether the needs and objectives of the user have been met. Thus, each time the life cycle passes through this phase, evaluation involves



**FIGURE 13.1** Which is the most appropriate way to evaluate different types of human-computer interaction?

\* ISO 13407 (1999) has largely been replaced by the revised ISO 9241 (2010) although it remains an interesting document in its own right.



**FIGURE 13.2** Human-centred systems design life cycle. (From ISO 13407, Human-centred design processes for interactive systems, International Standards Office, Geneva, Switzerland, 1999 [superseded by ISO 9241-210 (2010)].)

some form of measurement of the attributes of the design concept or product in order to make a judgement as to whether or not it is acceptable to users.

A point to note from this notion of stages in design life cycles is that evaluation plays different roles at each stage. This means, first and foremost, that evaluation is not a one-off activity to be conducted at the end of the design life cycle in order to allow a design to be signed off. Rather it means the following:

1. Evaluation is a recursive activity that cuts across the entire design life cycle.
2. Evaluation should be incorporated into as many stages of design as possible.
3. Evaluation should be designed to maximise the impact of the evaluation of the design stage in which it is used.
4. Evaluation should guide and inform design activity.

Methods that are applied during the design life cycle are called ‘formative’ (and directed at improving, modifying or otherwise adapting the design), while those at the end of the design life cycle are ‘summative’ (and directed at determining the fitness for the purpose of the product). This distinction between ‘formative’ and ‘summative’ implies different goals in conducting evaluation, and this is important to note. The point of ‘summative’ evaluation is really to test the fitness-for-purpose of the design (where fitness is defined by whatever acceptance criteria one is using) and is that point in the design life cycle when it is possible to conclude that the process is complete and the design is ‘accepted’. Obviously, there are many people (myself included) who would question whether design can ever be called ‘complete’ (because there are always modifications, improvements and alterations that can be made to the design, especially as it becomes used in contexts beyond those for which it was originally conceived). This means that there will be situations in which the summative evaluation of one design could initiate the process of new designs (and, in a sense, act as a formative evaluation in an extended project). The point of ‘formative design, on the other hand, is much less to do with metrics and measurement and much more to do with gaining insight into the real challenges that the context is presenting. In other words, having users engage with the design (in whatever form that one is able to present it to them) can have substantial impacts on the understanding of the design team, for example in terms of the needs of the end users, the types of tasks which

need to be supported and the effects that the context might have on defining performance. If I were to summarise advice on evaluation, it would be to make effort to conduct 'formative' evaluation as often as possible, with the intention of using this to refine and develop one's understanding of what is being designed, and then to conduct 'summative' evaluation as rigorously as possible, using measures in which one has confidence and which will reflect the criteria against which their design will be judged.

As well as these considerations of when to conduct evaluation, it is also important to consider *where* to conduct evaluation. This is an issue which has become increasingly important with the advent of personal, mobile and wearable technologies. In the past, evaluation was something which was most likely to be performed in a laboratory setting (i.e. a usability lab), and most large organisations have such settings to support their development work. In some instances, these labs are similar to the conventional computer software development laboratory, but more likely they are set out and equipped to provide a setting in which evaluators are able to perform tasks in a setting which feels as much like the environment in which the product will be used as possible. For example, in tests of home automation, the lab might be laid out like a person's lounge with television, audio and other technology with which they will be asked to interact. However, the usability lab provides a range of opportunities for analysts to collect data on their performance, for example close-circuit television or other forms of camera directed at the evaluator(s), mirrored walls to allow discrete observation and logging of keystrokes and mouse activity from the test computer.

In contrast with the lab-based approach, evaluation in 'real' settings can be performed in a person's workplace. This provides an opportunity for the day-to-day interruptions and distractions to activity and the range of other equipment and information sources that are typically used to be introduced into the evaluation. The obvious trade-off between the lab and the real-work environment relates to the degree of control that the analyst has over the performance of tasks and the collection of data *versus* the ecological validity of the evaluators' performance. It also has an impact on the type of data which can be collected. While the usability lab can capture quantitative data relating to interaction with the product (either through logging of performance or asking the users to 'talk aloud' as they use the product), the workplace will tend to limit the collection of such data and so will typically result in more qualitative data through observation or interview. For some analysts, the artificiality of the laboratory setting means that one is never going to gain insight into real activity. The extent to which this is an issue depends, I feel, on the type of product being evaluated and the range of tasks under test. Furthermore, there is an increasing range of products that can be used to monitor and record human activity in the workplace (ranging from 'apps' on smart phones which can log activity, location, etc., to monitoring software running on computer in offices, to sensors which can be placed on the person to record their activity). This range of technology, in effect, extends the usability lab into the workplace. It is worth noting that, just because it is possible to collect data in work environments, this does not mean that such data collection is not without problems. First, it is vitally important to ensure that the ethical considerations (in terms of informed consent) are applied to this non-laboratory setting. Second, the laboratory offers opportunities to control extraneous variables (such as interruptions from phone calls or work colleagues), which are far less easy to manage in the workplace. This means that one might collect the same form of data in the workplace as one collects in the laboratory, say in terms of keystrokes, but care must be taken with their interpretation. The workplace, therefore, has the potential to be a less sanitised environment than the laboratory. However, the trade-off between the control of variables in the workplace is the opportunity to study 'real' behaviour with the technology as opposed to the 'artificial' behaviour that one sees in the laboratory. Such ecological validity is a commonly cited benefit of studying activity in the environment in which it is typically performed. When it comes to the study of 'mobile devices', Kjeldskov and Graham (2003) noted that the majority of studies reported in the literature they reviewed (79%) involved laboratory studies, and only 19% involved fieldwork. However, in a subsequent research, Kjeldso et al. (2004) observed that field testing of mobile devices offered little additional benefit, in terms of identifying usability problems or workload considerations,

over conducting the research in laboratory settings (with the added concern that the ‘field’ settings (in this case hospital wards) led to loss of control of the variables affecting performance). Overall, the evaluation identified 37 problems (36 of the 37 problems were experienced by participants in the laboratory study and 23 of the 37 problems were experienced by participants in the field study) and that the field studies took 65 hours in comparison with 34 h for the laboratory studies. The suggestion is that field studies can be more time consuming with little obvious return in terms of the results produced. Rather than diverting into the well-trodden ‘laboratory’ versus ‘field’ study argument, I think it is very important to note that the ‘laboratory’ in which Kjeldsov et al. (2004) undertook their study was a carefully designed simulation of the working environment. Thus, the point at issue is less to do with the broad setting per se and much more to do with the key features of that setting.

This discussion re-emphasises the point that different products might involve the performance of different tasks in different environments under different conditions. It also suggests some difference in the approach that might be taken to defining ‘performance’.

## DEFINING ‘USABILITY’ AND ‘USER EXPERIENCE’

The term usability is generally agreed to have been defined by Shackel (1984), who referred to ‘...the capability in human functional terms [of a product] to be used easily and effectively by the range of users, given specified training and user support, to fulfil the specified range of tasks, within the specified range of environmental scenarios’ (Shackel, 1984). This definition has been developed into several international standards. For example, ISO 9241-11 (1998) defines usability as ‘...the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.’ According to ISO 9216 (2000), usability comprises ‘...a set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by stated or implied set of users’ and ‘...the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions’. ISO 9216 (2000) covers the apparent discrepancy between these definitions (i.e. is usability an attribute of the product or a consequence of using the product?) by defining a further concept, ‘Quality in use: the capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified contexts of use’. Table 13.1 illustrates the main components of ISO 9126’s (2000) concept of ‘quality in use’.

There are two points of note from this brief consideration: (1) each definition of usability, implicitly or explicitly, emphasises the notion of ‘context of use’ as a key component of usability and (2) usability can be measured in terms of at least 26 factors (see Table 13.1). Thus, ‘usability’ ought to read as a shorthand description of the complex interrelationship between people and technology, and most definitely *not* as an attribute of a product. Usability is often measured in terms of metrics. Following the lead of ISO 9241 (1998), these metrics can be grouped under the general headings of efficiency, effectiveness and satisfaction. The first two (efficiency and effectiveness) are often difficult to disentangle (because one tends to imply the other). Broadly, effectiveness concerns the ability of the user to achieve the goals set in the usability evaluation. This can be measured in terms of the number of tasks completed (assuming that the goal can be decomposed into discrete tasks) or in terms of the percentage of each task correctly completed, which implies some measure of error (although some authorities offer ‘error’ as a separate measure, it makes sense to include it under the heading of effectiveness). In contrast, efficiency concerns the effort required by the user to achieve the goals and can be measured in terms of completion time (either to achieve the overall goal or in terms of time to complete each task). These metrics have a strong quantitative basis and assume that activity has countable aspects (although there are also qualitative measures of effectiveness and efficiency, e.g. in terms of questionnaires, user reports and cognitive walk-throughs elicited from users during or after task performance).

What is apparent from this discussion is that usability often has a particular focus on the notion of ‘performance’ of users of products. Greenberg and Buxton (2008) argue that there are situations

**TABLE 13.1**  
**Quality in Use**

**Effectiveness, Productivity, Safety, Satisfaction**

|                        |                      |
|------------------------|----------------------|
| <i>Functionality</i>   | <i>Reliability</i>   |
| Accuracy               | Maturity             |
| Suitability            | Fault tolerance      |
| Interoperability       | Recoverability       |
| Security               | Availability         |
| <i>Usability</i>       | <i>Efficiency</i>    |
| Understandability      | Time-behaviour       |
| Learnability           | Resource utilisation |
| Operability            |                      |
| Attractiveness         |                      |
| <i>Maintainability</i> | <i>Portability</i>   |
| Analysability          | Adaptability         |
| Changeability          | Installability       |
| Stability              | Co-existence         |
| Testability            | Replaceability       |

*Source:* ISO 9126, Software engineering – product quality, International Standards Office, Geneva, Switzerland, 2000.

in which it is inappropriate to impose a usability evaluation on a product. However, rather than seeing this as a call to drop ‘evaluation’, this is a call to consider other ways to judge ‘worthiness’. The point is that the evaluation of usability is more appropriate in some stages of the design cycle than others, particularly in the notion of quantifiable measures of user performance.

The notion of usability has always turned to the somewhat tricky notion of ‘satisfaction’, but often struggled with its articulation (beyond simply seeing it as the subjective response of users to either the product or their performance). Looking at the struggles that usability measurement has had with this term, it is not surprising that HCI has taken a ‘turn to experience’ (Wright and Blythe, 2007). One can see that the notion of user experience (UX) is a broader, more holistic perspective on ‘satisfaction’ (Hassenzahl and Tractinsky, 2006). While usability tends to focus (but not exclusively on) on the relationship between a specific type of user performing a specific task in a specific environment, the scope of UX is much larger. Vermeeren et al. (2010) point out that UX is not only something which can be evaluated after a person has used a product or during this interaction, but also *before* that interaction. While the notion of a person’s perceptions, expectations and attitudes towards a product prior to interaction could conceivably be incorporated into the notions of usability, it is rarely something which plays a role in such evaluation. However, the values that a person brings to their experience of a product could have a bearing on how they may respond to it (Kujala and Väänänen-Vainio-Mattila, 2009). ISO 9241-210 (2008), which supersedes ISO 13407 (1998), defines UX as ‘...all aspects of the user’s experience when interacting with the product, service, environment or facility’. In reality, UX evaluation focuses on definable aspects, for example design aesthetics, pleasure in using the product, hedonic goals as well as the overall quality of experience arising from the interaction. For some authors, this distinction is sufficient to imply that the productivity-focused approach of usability is being superseded by the quality-focused approach of UX. Such a distinction blurs the benefit to be gained from each approach, and this chapter will treat these approaches as siblings: they share some family resemblances but have their own quirks and foibles.



## WHAT DOES HCI EVALUATION INVOLVE?

In order to consider the role that evaluation plays in HCI, it is useful to begin with some basic assertions. While these assertions might apply to a range of domains, the following discussion is couched in terms of evaluation in HCI. The assertions are as follows:

- Evaluation is integral to the process of design – we have considered the relationship between evaluation and the design life cycle given earlier.
- Evaluation measures more than a product's features – the preceding discussion has highlighted that evaluation involves the user's response, the manner in which they use the product.
- Evaluation implies critique.
- Evaluation involves measurement.
- Evaluation requires comparison with a referent model.

### EVALUATION IMPLIES CRITIQUE

Any form of evaluation implies a critique of the thing being evaluated. This critique might simply involve a potential user (or, indeed, a member of the design team) deciding whether she likes a particular product or whether she wishes to purchase it, or it might involve asking how might this thing be improved. The role of critique in evaluation is very important because it provides a primary motivation for the type of measurement which might need to be applied. One can imagine situations in which the evaluation is mistakenly assumed to be simply a matter of expressing a personal opinion. My own experience of teaching HCI evaluation suggests that this is not an unusual first reaction to the challenges of conducting evaluation. It offers a number of attractions, not least of which is the lack of requirement to provide any evidence in support of an opinion once it has been asserted, that is simply saying 'I like that' or 'I don't like this' provides little useful information. However, even faced with an approach to evaluation which takes such a solipsistic stance, it is possible to probe further and more clearly establish a form of critique. For example, one could apply techniques such as the '5 Whys' (originally developed in the Toyota Motor Corporation to provide a basis for critiquing manufacturing processes). In this technique, one simply keeps asking 'why?' to force expansion of a statement. For example, imagine a situation in which an 'evaluator' (say a teenager) is presented with a set of products (say new designs of smart phones) and asked to pick their favourite. They pick product X and the first 'Why?' elicits the response, 'Because it's cool'; the second 'Why?' results in 'Its graphics are neat'; the third 'Why?' produces 'They're really clear and bright and the screen's a good size'; the fourth 'Why?' could be phrased as 'Why is this important?', leading to 'So I can watch movies and play games'; the fifth 'Why?' could lead to the conclusion that 'This is the best phone to use when I'm travelling to college'. In this example, the 'evaluation' (while based on the opinions of an individual) provides some 'evidence' which could be presented to other people to see if it produces consensus. This also highlights the problem of using a single sample and an unstructured approach – asking a second person could produce different choice based on different opinions. However, it does indicate the ways in which the role of critique can be used to explore people's reactions to technology. This example also highlights the problems of controlling variables in evaluation. In this example, different people might present different opinions and so achieving consensus might be problematic. The simplest solution to these problems is to make sure that several methods are used to support the evaluation and that some form of triangulation is then performed across these methods, that is cross-checking the results of one method with those of the others.

### EVALUATION INVOLVES MEASUREMENT

Nielsen and Levy (1994) pointed out that usability itself cannot be measured, but aspects of usability can. The International Standards, discussed earlier, specify the need to measure aspects of HCI evaluation and provide an indication of what to measure, but they leave the precise definition to

the evaluator. This is primarily because, unlike physical measurements such as length or weight or voltage, the application and interpretation of the measure will vary according to the context of use. This inevitably makes the idea of a standard measure of usability or UX highly problematic. Whilst one can define the measurement test and report results, it is not clear how one ought to compare the outcome of one test with another (particularly when slight variations in test might be used). For example, can one say that system X (for remote monitoring of asthma patients) which scores '3' is, therefore, inferior to system Y (for computer-aided design) which scores '7'? Having said this, if one cannot report an agreed set of benchmark measures, how can one speak of measurement? It is like using the notion of a 'hand' to measure the height of a horse, allowing both children and adults to use their own hands as reference and then wishing to argue that a pony is bigger than a carthorse. Not only does this raise the problem of assuming that the same 'scale' can be applied to different contexts to produce comparable results, it also misses the ways in which the interpretation of these results are context dependent. Thus, it is often useful to not only identify a usability 'score' but also to indicate the 'risk' that failing to meet this usability score could produce. In this manner, the scoring of usability becomes a matter of scoring particular dimensions of usability as, e.g., produced by the Software Usability Scale (SUS) of Brooke (1996) in terms of their perceived likelihood and perceived impact, or severity, on completing a given task.

Springett (2008) draws a distinction between measurement as quantification and proof on the one hand, and measurement as assessment against criteria or towards a goal on the other hand. I suggest that the notion of measurement in HCI is primarily directed at defining consensus. This consensus could be achieved through the collation of responses from several users or through the collection of data through a standard methodology. Consensus would help provide confidence in the outcome of the evaluation. This gives a clue as to the genesis of many of the approaches employed in HCI evaluation: the idea that the outcome of an evaluation could be generalised with confidence lies at the heart of many of the methods that are employed. Such an idea has its roots in laboratory experiments. To some extent, the methods are variations of techniques and methods reviewed elsewhere in this book. However, the focus of the evaluation is primarily directed at the design of the product under consideration (as opposed to measuring user performance *per se*). This raises a different idea as to the nature of evaluation. Furthermore, the very notions of persuasion or evidence (and implied notions of consensus that they embody) can be difficult to apply to some of the fundamental aspects of UX. For example, if one considers the aesthetics of a designed object to be reflected in the emotional response of the people who encounter it, does it make sense to assume that they will share the same emotion (or at least be persuaded into a consensus), or is it more likely that some people will respond positively while others might be less positive?

Standards relating to usability seek to operationalise the components of usability and to define benchmarks against which the product could be tested (although these standards do leave the definition of many of the components, measures and tests fairly open ended). In many respects, this is similar to the manner in which concepts such as 'quality' are handled – the evaluator sets out a set of terms of reference, providing specific definitions of the terms, and then seeks to measure against this set.

## EVALUATION REQUIRES COMPARISON WITH A REFERENT MODEL

However one seeks to measure a product, the measurement will make sense only in terms of some point of reference. In orthodox approaches to measurement, this could take the form of an agreed scale. In evaluation, this could equally take the form of a competitor product or the opinions of other people. Any form of evaluation is an 'evaluation against' something; evaluation requires a referent model. It is naive to believe that one can 'evaluate' something in a vacuum, that is to think that one can take a single product and 'evaluate' it only in terms of itself. At best, the referent model would be the assumptions and preconceptions of the design team; at worse, the preconceptions might not be articulated, and the evaluation becomes difficult to interpret. If you do not clearly state your assumptions and define the 'referent model' for your evaluation, then any conclusions are meaningless.

### Evaluation against Other Products

One of the problems with the view that ‘evaluation against other products’ ought to focus on ‘products’ *per se* is that one rarely finds a set of products that can be matched on all functions, for example one might find that product X is excellent for performing activity j and very poor for activity i, whereas the converse is true for product Y, and that product Z does not even support activity j. To make matters worse, one might not be able to find any products that are comparable with the new product (you might have such a radically new design that there is nothing like it in existence). The question for ‘evaluation against other products’, therefore, hinges on the issues of what functions to address, what aspects of performance to measure and how to make sensible comparisons. Having said this, producing an initial table summarising whether the packages offer specific functions can be informative and can highlight initial differences between them. Such a listing of functions offered by packages represents a simple form of comparison and one that can be used to discuss the pros and cons of the packages in general terms, but this might only serve to highlight simply that the items that one wishes to compare are not identical. Obvious or not, one is typically comparing different things in HCI evaluation. It is my proposal that conducting evaluation against other products in terms of a set of activities offers the analyst the following benefits:

- The evaluation will cover a range of functions on the products. It is important to ensure that the comparison provides a fair and accurate view of the product. After all, it is not really the point of evaluation to just demonstrate that product X is better than product Y – partly because there are bound to be occasions when products X and Y are similar, or where product Y is better than product X, and partly because simply knowing that  $X > Y$  tells us very little about how to improve X (or Y) or why X is superior.
- The focus of the evaluation is less on product functioning than on user activity. This might appear, at first glance, to be tautological – surely, product evaluation is about evaluating the product. HCI is about human-computer interaction, and the defining feature of this relationship is the interaction (rather than either human or computer).
- When comparing user activity on two or more products, it is important to decide what information is really being sought. Of course, it depends what one thinks one needs to know: however, I would suggest that conclusions are more useful if they are not merely dependent on that task being performed by those participants with those devices under those conditions, but allow the designer to reason about possible variations in performance and to generalise to new situations and new designs.

### Evaluation against Design Targets

The notion of evaluating against design targets is central to the potential role of evaluation as a significant component of the human-centred design life cycle. There are several reasons why design targets are beneficial. First, the design team can specify the type of targets that they believe are important, for example if the technology is intended to support training, the design team will need to specify a target relating to training effectiveness. Second, the design team can set increasingly specific targets as a means of supporting rapid prototyping, that is when a version of the prototype meets one set of targets, the team can decide whether to change the target. Third, the design team will be able to demonstrate performance improvements through showing either how the targets have changed during the design process or showing how the product has been modified to meet the targets. Fourth, the International Standards (discussed previously and later) can be easily related to design targets; for example under the ISO 9241 (1998) notion of usability, one could begin with a target of ‘66% of the specified users would be able to use the 10 main functions of product X after a 30 minutes introduction.’ Once this target has been met, the design team might want to increase one of the variables, for example 85% of the specified users, or 20 main functions or 15 minutes introduction.

The notion of specifying targets lends itself to the development of usability specifications. In an early discussion of usability engineering, Good et al. (1986) propose that it is important to define

both usability goals and metrics that relate to these goals. Given these notions, it is then possible to determine planned or acceptable levels of performance on each metric. For example, in a study of conferencing systems, Whiteside et al. (1988) identified 10 attributes that they felt reflected the use of the conferencing system, for example ranging from a fear of feeling foolish to user preference, to a number of errors made during task performance. For each attribute, Whiteside et al. (1988) defined a method for collecting data about that attribute, for example questionnaires and observation and then set performance limits relating to best, worst and planned levels. It is not always easy to quantify design targets. However, it is possible to advance a set of qualitative measures that could be considered in evaluation. For instance, Holcomb and Tharp (1991) propose a 'model' of interface usability that contains seven components: functional, consistent, natural and intuitive, minimal memorisation, feedback, user help and user control.

### **Evaluation against User and Organisational Requirements**

HCI has long recognised that a core feature of design is the capturing of user requirements and the derivation of specifications from such requirements. It has long been recognised that such an approach is near impossible, and only the naive and inexperienced believe that this can be done. There are three problems with this view: (1) users do not know what they want and have great difficulty in expressing these wants (Norman, 1988). This is not to say that users cannot contribute enormously to the design process. Indeed, some of the participative design and evaluation methods (see Chapter 34) demonstrate that user involvement can lead to many benefits to both the design team and to the users themselves; (2) user requirements are typically incomplete, grounded in everyday experience and do not reflect the future system. This means that any specification taken solely from user requirements will be flawed. This is not to say that users should be left out of the requirement process, but that simply asking them is not enough; (3) any specification (typically) represents a 'model' of the future system, that is a formal description of what the system will look like and how it will function. Any model, by definition, is a simplification of a complex process and (probably) omits significant aspects, either to make the model tractable (i.e. easy to develop) or to ensure coherence (i.e. to avoid contradictions in requirements – even when these exist) or to fit the modelling language (i.e. a dataflow diagram does not easily capture the interpersonal politics between members of a management team, but such 'politics' might affect how the system is used).

### **Evaluation against Standards**

When one thinks about standards in engineering, one probably assumes that these documents define ranges of performance that products or devices should meet in order comply with the standards or define procedures that one needs to follow in order to demonstrate compliance. While there are some standards that cover physical aspects, such as key size or displacement, the more recent standards for HCI are somewhat different in that they tend not to report quantitative limits that products need to meet. There is a simple reason for this; as mentioned earlier, HCI is not merely about studying products (or users) but about the interaction between users and computers as well. This means that it is meaningless to define absolute limits of performance that all devices, products or systems would need to reach in order to meet standards. For example, would it be sensible to require that a computer keyboard should always result in typing speeds of 100 words per minute, that is irrespective of user capabilities, task demands, etc.?

As Bevan (2001) points out, HCI standards are generally concerned with the following areas:

1. The use of the product
2. The user interface and interaction
3. The process used to develop the product
4. The capability of an organisation to apply user-centred design

The focus of the standards being developed by the International Standards Organisation's (ISO) Technical Committee ISO/TC 159/SC4 Ergonomics of human–system interaction has been to develop documentation, procedures and support to enable people developing interactive systems to design and develop products that are demonstrably able to support effective, efficient and satisfying performance by specified users in specified contexts. For example, ISO 9241 (1998) offers procedures on conducting evaluations, for example, for measuring glare and reflection from VDTs (ISO 9241-7), for assessing interaction devices (ISO 9241-9) as well as guidance on specific aspects of HCI, for example use of colour on displays. ISO 13407 (1999) is concerned with encouraging the involvement of human-centred design processes and activities throughout the life cycle of interactive systems. The primary characteristics of the human-centred approach advocated by ISO 13407 (1999) are

- a. The active involvement of users and a clear understanding of user and task
- b. Requirements
- c. An appropriate allocation of function between users and technology
- d. The iteration of design solutions
- e. Multidisciplinary design

ISO13407 advises the use of an evaluation plan (7.5.2) and suggests that the evaluation should provide feedback into the design process (7.5.3), using either experts or users for this process.

## AN OVERVIEW OF USABILITY EVALUATION METHODS

Throughout the 1990s, there were attempts to compare usability evaluation methods (Jeffries et al., 1991; Nielsen and Levy, 1994; Frokajer et al., 2000). Gray and Salzman (1998) presented a damning indictment of five studies comparing evaluation methods, particularly in terms of the small sample sizes used in the comparisons, the lack of control of variables and inappropriate reporting and manipulation of data. Andre et al. (1999) consider 17 usability evaluation studies and note a general lack of consistency between findings. They also criticise the methodology followed in the comparison studies. Thus, the question of which methods are 'best' remains open. This is a common problem in ergonomics research (Stanton and Young, 1999), and there remains a tendency to employ methods that 'feel' appropriate rather than investigating their appropriateness.

### ANALYTIC METHODS

Analytic methods can be used to predict the likely performance of users with a particular device (Card et al., 1983; Olson and Olson, 1990) or error (Young and Whittington, 1990; Baber and Stanton, 2001). An example of such an approach is the keystroke-level model (Table 13.2) in which a set of tasks are defined in terms of basic times, and these tasks are combined to produce a predicted time to complete the sequence of tasks (Card et al., 1983).

One use of such models would be to define performance benchmarks; for example one could predict that the design of device X ought to support a performance time of 3 s to complete a specific task. One could then use this target performance time as a criterion for performance testing, for example stating that you expect, say, 66% of your user group to complete the task in 3 s (or less). A second use of these models is to conduct 'evaluation' of products prior to building. For example, Gray et al. (1993), using critical path analysis, were able to demonstrate that a proposed computer workstation to support call handling (for a major telephone company) would not affect overall call handling time as the use of the computer did not lie on the critical path (most of the operator's time was spent speaking to the caller, rather than using the computer). Consequently, the study showed that investment in the technology would have resulted in an undue expense and not net gain in productivity.

**TABLE 13.2**  
**Brief Description of KLM**

|                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------|----------|-------------------------------------|--|------------------------|--|------------------------|----------|-------|--------|-------|------------------|--------|
| Name of Method:      | Keystroke-Level Model (KLM)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Evaluation type:     | Analytic                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Design life cycle: Pre-design; early           |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Background knowledge | HCI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Brief description    | <p>The method employs a form of task analysis, in which a goal is decomposed into subgoals and tasks. The subgoals and tasks can then be related to standard description, using either production rules or using standard times, such as follows:</p> <table> <tr> <td>Keypress</td> <td>0.12 s (90 words per minute typist)</td> </tr> <tr> <td></td> <td>0.5 s (random letters)</td> </tr> <tr> <td></td> <td>0.75 s (complex codes)</td> </tr> <tr> <td>Pointing</td> <td>1.1 s</td> </tr> <tr> <td>Homing</td> <td>0.4 s</td> </tr> <tr> <td>Mental operation</td> <td>1.35 s</td> </tr> </table> |                                                | Keypress | 0.12 s (90 words per minute typist) |  | 0.5 s (random letters) |  | 0.75 s (complex codes) | Pointing | 1.1 s | Homing | 0.4 s | Mental operation | 1.35 s |
| Keypress             | 0.12 s (90 words per minute typist)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
|                      | 0.5 s (random letters)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
|                      | 0.75 s (complex codes)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Pointing             | 1.1 s                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Homing               | 0.4 s                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Mental operation     | 1.35 s                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Referent model       | <p>KLM, in common with the various other analytic techniques, have two primary reference models:</p> <ol style="list-style-type: none"> <li>1. The models' ability to predict user performance is assessed through comparison with data from users performing the tasks.</li> <li>2. The models are aimed at developing their own referent models, employing standard descriptions or performance data. The validity of the models can be open to question because it is not straightforward to compare the prediction against other data.</li> </ol>                                                 |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Source references    | <ol style="list-style-type: none"> <li>1. Card et al. (1986)</li> <li>2. Olson and Olson (1990)</li> <li>3. Stanton and Young (2001)</li> </ol>                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Performance          | Number of evaluators                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 1 <sup>[1,2]</sup>                             |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
|                      | Duration                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Minutes to hours (depends on analyst and task) |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
|                      | Agreement with user data                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 80%–90% <sup>[1]</sup>                         |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
|                      | Inter-rater reliability                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | d = 0.754 <sup>[3]</sup>                       |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |
| Overall              | <p>Reducing human activity to standard tasks. Can be useful to develop predictions of performance. However, such a reductionist approach might omit key features of performance or ignore the influence of contextual features.</p>                                                                                                                                                                                                                                                                                                                                                                   |                                                |          |                                     |  |                        |  |                        |          |       |        |       |                  |        |

## USABILITY INSPECTION METHODS

Usability inspection concerns the evaluation of products through the examination of their features. Typically, the evaluation considers the 'static' aspects of the features, that is the inspection will be performed on the features that are shown on the interface. In the 1990s, there was a call for usability inspection and evaluation methods to be 'quick and dirty'. This led to various forms of 'discount usability' techniques, of which Nielsen's (1993) various accounts of heuristic evaluation are, perhaps, the best known. These methods could be performed with minimal outlay, experience, effort or users, and could produce some useful data. In a provocative statement at *CHI'95*, Gray et al. (1995) argued that usability inspection methods are of dubious utility. They noted that HCI has turned away from the idea of evaluating against guidelines (e.g. Smith and Mosier, 1986). The primary reasons why HCI guidelines fell out of favour was that they were tremendously tedious to apply and often failed to capture the right level of detail, that is the guidelines were either too prescriptive (and so could not easily be applied to all designs) or too vague (and so could not be applied with any rigour). However, at least the guidelines represented an agreed referent model, that is a 'good' interface was one that complied with the guidelines. Tables 13.3 and 13.4 show two commonly employed usability inspection methods: heuristic evaluation and cognitive walk-through.



**TABLE 13.3**  
**Brief Description of Heuristic Evaluation**

|                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                         |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|
| Name of Method:      | Heuristic Evaluation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                         |
| Evaluation type:     | Inspection                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Design life cycle: During use           |
| Background knowledge | Product, task or HCI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                         |
| Brief description    | <p>The evaluator is asked to consider the product in terms of a set of simple rules (see later). For each rule, the evaluator determines how the product meets (or fails to meet) the rule.</p> <ol style="list-style-type: none"> <li>1. Use simple and natural language</li> <li>2. Provide clearly marked exits</li> <li>3. Speak the user's language</li> <li>4. Provide short cuts</li> <li>5. Minimise user's memory load</li> <li>6. Good error messages</li> <li>7. Be consistent</li> <li>8. Seek to minimise user error</li> <li>9. Provide feedback</li> </ol>                                                                           |                                         |
| Referent model       | <p>With the heuristics listed earlier, the degree to which a product meets a rule could be indicated by a rating scale. This provides an indication of where the product is more or less successful in meeting these heuristics (and, thus supports the role of critique). However, a rating by itself does not tell you whether the score is good, poor or average – for this, one needs to have some experience of conducting evaluation of other products. For example, a useful exercise is to apply these heuristics to well-known commercial products and then use this as the benchmark to consider for the product under consideration.</p> |                                         |
| Source references    | <ol style="list-style-type: none"> <li>1. Nielsen (1993)</li> <li>2. Andre et al. (1999)</li> <li>3. Hertzum and Jacobsen (2001)</li> <li>4. Spool and Schroeder (2001)</li> </ol>                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                         |
| Performance          | Number of evaluators                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 3–5 <sup>[1]</sup> ; >15 <sup>[4]</sup> |
|                      | Duration                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Minutes                                 |
|                      | Inter-rater agreement                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 5%–65% <sup>[3]</sup>                   |
|                      | Problems found                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 17%–46% <sup>[2]</sup>                  |
|                      | Agreement with 'expert'                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 20%–86% <sup>[2]</sup>                  |
| Overall              | <p>Quick method to apply and so can be useful for rapid evaluation of designs. However, the method is highly subjective and depends on the experience of the evaluator and interpretation of the rules, so open to bias. Furthermore, the method is dependent on the 'detectability' of the problem.</p>                                                                                                                                                                                                                                                                                                                                            |                                         |

## USER REPORTS

User reports are employed by all three perspectives, although a craft or engineering approach might use less structured methods than the applied science approach. It is a truism to state that '...ultimately it is the users of a software system [or any product] who decide how easy its user interface is to manipulate...' (Holcomb and Tharp, 1991, p. 49). One might feel that asking people about the product would be the obvious and most useful approach to take. At a craft level, this might simply mean using a fairly open and unstructured interview to elicit user opinion (see Chapter 4). While such an approach might yield some useful insight, it can lack the potential to provide generalisable data. One of the problems that is worth noting before discussing some of these techniques is that there are few 'off-the-shelf' methods that can be employed. Consequently, evaluators will need to develop their own methods and instruments. This can be time-consuming (particularly if one is planning to use a questionnaire). However, it is possible that companies simply reuse existing instruments.

There are many surveys that have been designed to evaluate usability. Some, like CUSI – Computer User Satisfaction Inventory (Kirakowski and Corbett, 1988) and QUIS – Questionnaire for

**TABLE 13.4**  
**Brief Description of Walk-Through**

|                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
|------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|---------------------|----------|-----------------------------------------------|-----------------------|-----------------------|----------------|-----------------------|-------------------------|-----------------------|
| Name of Method: Walk-Through |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Evaluation type: Inspection  | Design life cycle: During use                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Background knowledge         | Product, Task or HCI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Brief description            | <p>The evaluator is asked to perform specific tasks, or to achieve specific goals, using the product and to provide a simultaneous commentary. The commentary should contain what the evaluator is thinking, doing or looking at. The technique is a form of verbal protocol (see Chapter 8). The resulting transcript can be considered in terms of the following questions:</p> <ol style="list-style-type: none"> <li>1. What does the user want to achieve?</li> <li>2. Are the correct actions sufficiently clear for the user?</li> <li>3. Will the user connect the correct action with the system image?</li> <li>4. Will the user correctly interpret the system response?</li> </ol> |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Referent model               | At heart, the referent model for walk-through is an (assumed) error- or problem-free performance of the task. It is often beneficial for the analyst to either perform the selected tasks prior to asking the evaluator to attempt it or ask a product-expert to perform the task as a walk-through. This helps to provide an indication of what ‘good’ activity might look like and frames the walk-through as a process to identify areas where activity could become ‘poor’, for example as a result of problems in the product’s design.                                                                                                                                                   |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Source references            | <ol style="list-style-type: none"> <li>1. Newman and Lamming (1995)</li> <li>2. Andre et al. (1999)</li> <li>3. Hertzum and Jacobsen (2001)</li> </ol>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Performance                  | <table> <tr> <td>Number of evaluators</td><td>2+<sup>[1,2]</sup></td></tr> <tr> <td>Duration</td><td>Minutes (depends on task and detail required)</td></tr> <tr> <td>Inter-rater agreement</td><td>5%–65% <sup>[3]</sup></td></tr> <tr> <td>Problems found</td><td>7%–17% <sup>[2]</sup></td></tr> <tr> <td>Agreement with ‘expert’</td><td>5%–73% <sup>[2]</sup></td></tr> </table>                                                                                                                                                                                                                                                                                                          | Number of evaluators | 2+ <sup>[1,2]</sup> | Duration | Minutes (depends on task and detail required) | Inter-rater agreement | 5%–65% <sup>[3]</sup> | Problems found | 7%–17% <sup>[2]</sup> | Agreement with ‘expert’ | 5%–73% <sup>[2]</sup> |
| Number of evaluators         | 2+ <sup>[1,2]</sup>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Duration                     | Minutes (depends on task and detail required)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Inter-rater agreement        | 5%–65% <sup>[3]</sup>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Problems found               | 7%–17% <sup>[2]</sup>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Agreement with ‘expert’      | 5%–73% <sup>[2]</sup>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |
| Overall                      | Method is relatively quick to apply and can provide useful insight into the misunderstandings that users might make when using the product. However, the results are highly dependent on the type of users employed and the type of tasks considered. Relatively quick method to apply, but can take time to transcribe or analyse results. However, the method depends on the evaluator’s ability to generate ‘problems’.                                                                                                                                                                                                                                                                     |                      |                     |          |                                               |                       |                       |                |                       |                         |                       |

User Interface Satisfaction (Chin et al., 1988), are designed to capture user response to an interface, particularly in terms of affective components (such as satisfaction). This approach can be seen as a simplification of the use of guidelines; the evaluator is provided with a set of characteristics of a ‘good’ interface and asked to state how well the product under evaluation compares. Some surveys, like Ravden and Johnson (1989), have been designed to cover both aspects of the interface and characteristics of usability. This checklist can also be used as source material for developing reduced evaluation sets. Other surveys, like SUS (Brooke, 1996), have been designed to support ‘quick and dirty’ evaluation. While the approach can be applied quickly and the items are relatively easy for the lay evaluator to interpret and apply, the use of 10 items could suggest that the approach is open to the same criticisms as heuristics. Table 13.5 shows SUMI – Software Usability Measurement Inventory (Kirakowski and Corbett, 1993).

## HCI IN THE WILD

The notion that studying human interaction with technology is best conducted through observing people actually using the technology is central to much of HCI research. In terms of evaluation, the goal of observation is often to see how technology affects user performance.

**TABLE 13.5****Brief Description of Software Usability Metric Inventory**

|                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Name of Method: Software Usability Metric Inventory (SUMI) |                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Evaluation type: User report                               | Design life cycle: After use                                                                                                                                                                                                                                                                                                                                                                                      |
| Background knowledge                                       | Product or task                                                                                                                                                                                                                                                                                                                                                                                                   |
| Brief description                                          | Developed as part of an EU project, SUMI provides a ‘benchmarked’ evaluation measure. Evaluators are asked to consider the product in terms of a 50-item survey. Each survey item is rated in terms of three responses (x, x, x). The evaluation requires a minimum of 15 evaluators to complete the survey.                                                                                                      |
| Referent model                                             | The results of the survey are fed into a computer and compared against a database of ‘standard’ scores. The database (derived from some 200 previous analyses) provides an indication of how well the software relates to other products and, as such, is useful in benchmarking exercises. The analysis produces scores relating to global usability, efficiency, affect, helpfulness, control and learnability. |
| Source references                                          | 1. Kirakowski and Corbett (1993)                                                                                                                                                                                                                                                                                                                                                                                  |
| Performance                                                | Number of evaluators 15+ <sup>[1]</sup>                                                                                                                                                                                                                                                                                                                                                                           |
|                                                            | Duration Minutes for each analysis, under 1 h for evaluation                                                                                                                                                                                                                                                                                                                                                      |
|                                                            | Agreement with other methods Produces similar findings to other user reports; internal verification, based on comparison with previous studies                                                                                                                                                                                                                                                                    |
| Overall                                                    | A 50-item questionnaire (available in English, Spanish, Italian, German and Dutch) which is completed by a group of evaluators after experiencing a product. The responses are compared with measures in a database of previous evaluations of representative applications and presented in terms of standardised scores.                                                                                         |

**PARTICIPATIVE EVALUATION**

One of the seminal works in the sociology of HCI is Suchman (1988). Central to this work was a study in which pairs of users of a complex Xerox photocopier were asked to perform a series of tasks. The conversations between the users were transcribed, and through conversation analysis, some of their misunderstandings and difficulties were illustrated and discussed. The idea of having pairs of users work together to examine products has grown in HCI and represents one of the common forms of participative evaluation. Recent developments of this approach involve pairs of users working together to use a product, either in terms of one person demonstrating the product to the other, or in terms of both the people cooperating in terms of using the product. Typically, the approach involves the pair being filmed and extracts of the film being presented to the design team in order to illustrate problems that are being experienced.

**OVERVIEW OF USER EXPERIENCE METHODS**

Usability often (but not always) focuses on the efficiency and effectiveness through which specific users perform specific tasks in specific environments, but it has less to say on the notion of ‘attitude’. This could be dealt with through the use of user surveys or questionnaires, but these often tend to focus on the relationship between the use of the product and the achievement of user goals, as opposed to more hedonic responses to the product itself. One of the differentiating features of UX is its focus on these hedonic aspects. As the methods considered in this section illustrate, capturing UX can be performed prior to encountering the product, or during use of the product or after the interaction/experience has finished. These latter two points to evaluate are, of course, also the province of usability, and it makes little sense to argue strongly for demarcation between the two traditions. Rather, the experienced evaluator will bring a collection of methods to bear on the challenge of evaluating a product; the choice of methods will depend on the nature of the product and the type of interaction which is of interest.

Bargas-Avila and Hornbaek (2011) reviewed over 60 studies in which UX is evaluated. They point out that the majority of the studies involve eliciting subjective response, particularly in terms

of emotional response to the product, enjoyment of the interaction and aesthetics of the product’s design. The paper notes the problems of validating many of these reports, particularly as so many studies rely on instruments developed by the authors of the various studies. A further issue relates to the various forms of ‘halo’ effect that one often has when people first encounter a novel product. There is a strong likelihood that positive response to the appearance of the product will influence other ratings, such as perceptions of usability, enjoyment and emotional response. However, such effects tend to diminish with repeated exposure, leading Sonderegger et al. (2012) to propose that evaluations ought to be spaced over a period of 2 weeks in order to counter such effect. UX evaluation is often, but not always, conducted over very short time periods. As such, it is far more likely to constitute formative evaluation, than summative evaluation, and is far more likely to be conducted to raise design issues which can be fixed during the subsequent development of the product (or which can form the basis of requirements for new designs) (Tables 13.6 to 13.8).

**TABLE 13.6**  
**Brief Description of Product Emotion Measurement Tool (PrEmo) PrEMO**

|                              |                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                        |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| Name of Method: PrEmo        | Design life cycle: Before use                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                        |
| Evaluation type: Self-report |                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                        |
| Background knowledge         |                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                        |
| Brief description            | Evaluators are given a set of 14 cartoon characters expressing different emotions and use these to indicate their response to a product. This gives a snapshot of some of the emotions, but the full set (and instructions for use can be obtained from <a href="http://studiolab.ide.tudelft.nl/studiolab/desmet/premo/">http://studiolab.ide.tudelft.nl/studiolab/desmet/premo/</a> ) |                                                                                                                                        |
|                              | Pleasant emotions                                                                                                                                                                                                                                                                                                                                                                       | Unpleasant emotions                                                                                                                    |
|                              |  Desire                                                                                                                                                                                                                                                                                               |  Boredom                                             |
|                              |  Fascination                                                                                                                                                                                                                                                                                         |  Contempt                                           |
|                              |  Pleasant surprise                                                                                                                                                                                                                                                                                   |  Dissatisfaction                                    |
|                              |  Satisfaction                                                                                                                                                                                                                                                                                        |  Disgust                                            |
|                              |  Amusement                                                                                                                                                                                                                                                                                           |  Unpleasant surprise                                |
| Referent model               | The data from the evaluators using the PrEmo software tool can be aggregated to produce an overview of the response to the product under test <sup>[1]</sup> .                                                                                                                                                                                                                          |                                                                                                                                        |
| Source references            | 1. Desmet et al. (2007)                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                        |
| Performance                  | Number of evaluators                                                                                                                                                                                                                                                                                                                                                                    | 10+                                                                                                                                    |
|                              | Duration                                                                                                                                                                                                                                                                                                                                                                                | Minutes                                                                                                                                |
|                              | Reliability                                                                                                                                                                                                                                                                                                                                                                             | The primary measure of reliability for this technique will be inter-rater reliability (or correlation of response amongst evaluators). |
| Overall                      | PrEmo provides users with a means of indicating their emotional response to a product. The collation of several responses into the analysis offers an empirical basis for contrasting responses, for example in terms of considering any groups that might differ. A cross-cultural comparison of interpretation of the pictures used to express emotion would be interesting.          |                                                                                                                                        |

**TABLE 13.7****Brief Description of Tracking Real User Experience (TRUE)**

Name of Method: Tracking Real User Experience (TRUE)

Evaluation type: User logging

Design life cycle: During use

Background knowledge

|                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                 |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Brief description | A combination of data logging and attitudinal data to reflect performance on playing video games. The data logging is aimed at capturing streams of data, rather than discrete events, which provides a good way of contextualising activity. The attitudinal data are collected at regular intervals when the game is paused, and the user is asked to respond to onscreen prompts.                                                                                                              |                                                                 |
| Referent model    | This approach has been used to study gameplay in a variety of video games. The collection of massive amounts of data provides the games designers with insight into how the games are being played. This suggests that the referent models are (a) the games designers' assumptions of how the game ought to be played and (b) the aggregated performance of many players defining particular types of gameplay (which can be compared against each other or against the designers' assumptions). |                                                                 |
| Source references | 1. Kim et al. (2008)                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                 |
| Performance       | Number of evaluators                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1–1000s <sup>[1-]</sup>                                         |
|                   | Duration                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Depends on question being asked, but could last an entire game. |
|                   | Agreement with other methods                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                 |
| Overall           | The use of some form of performance logging has a long tradition in usability evaluation, for example DRUM (Macleod and Rengger, 1993). However, TRUE offers an opportunity to collect richer, longitudinal data from video game players which represents a significant increase in both the range and the quantity of data collected and analysed.                                                                                                                                               |                                                                 |

**REPORTING EVALUATION**

While the same methods could conceivably be used for both formative and summative evaluation, experience suggests that the formative approaches tend to be applied in a manner which can allow rapid assessment of output, which can be fed directly into the ongoing design life cycle, where summative approaches tend to be applied in a manner which is amenable to publication in academic journals or technical reports. In terms of the type of reporting of evaluation that one sees in the literature (and formal reports), there is a tendency to favour summative evaluation. This may be a consequence of the scientific approach that is central to the conduct of summative evaluation, that is a representative sample of prospective users are selected and used to perform controlled tasks in a calibrated environment in order to provide data which can be amenable to statistical analysis. However, this represents only a fraction of the types of activity which can legitimately be termed evaluation in HCI.

The manner in which evaluation is reported will, of course, depend on the audience to which one is presenting. There might be a temptation to treat formative evaluation as informal, with minimal recording of findings and recommendations and an emphasis on sharing ideas amongst the design team in a way which will not interrupt to flow of the design activity. This is all well and good – until someone later questions a particular design decision. Thus, having some form of record is beneficial. However, this need not mean that the reporting has to involve the production of a detailed written report: for some forms of evaluation, one could video-record the evaluators conducting their test (perhaps having the analyst provide an introduction to the recording to highlight any key points); or the analyst could be videoed presenting the findings; or a high-quality digital photograph could be taken of the summary of findings presented during the design team meeting. The point is not that recording these ideas need be onerous, but that getting into the habit of capturing the recommendations can be useful, particularly in a large project which has many contributors. It is surprising how

**TABLE 13.8**  
**Brief Description of Experience Sampling Method (ESM)**

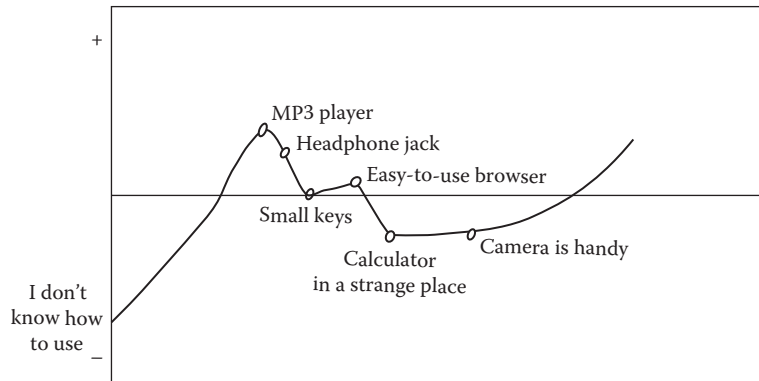
Name of Method: Experience Sampling Method (ESM)

Evaluation type: Self-report

Design life cycle: After use

Background knowledge

Brief description Following exposure to a product, evaluators are asked to indicate their changing experience of using a product by drawing a curve, for example



Short description of the changes:  
 Learning in the beginning

|                   |                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                   |
|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Referent model    | Following this, evaluators are asked to draw a further set of curves to indicate the attractiveness, ease of use, utility, degree of usage.                                                                                                                                                                                                                                               |                                                                                                                                                                                   |
|                   | The main reference for this technique appears to be the response of other evaluators. This allows the analyst to draw comparisons between evaluators with positive or negative responses, and trends in the evaluations over time. However, the use of the + and – partitioning of the curve indicates a further referent model relating to positive or negative response to the product. |                                                                                                                                                                                   |
| Source references | Kujala et al. (2011)                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                   |
| Performance       | Number of evaluators                                                                                                                                                                                                                                                                                                                                                                      | 1+                                                                                                                                                                                |
|                   | Duration                                                                                                                                                                                                                                                                                                                                                                                  | Minutes                                                                                                                                                                           |
|                   | Comparison with other methods                                                                                                                                                                                                                                                                                                                                                             | The approach could provide similar output to evaluations using semantic analysis (in that, evaluators are asked to consider the product against specific terms).                  |
|                   | Overall                                                                                                                                                                                                                                                                                                                                                                                   | The approach is intuitively appealing and simple to apply. Its strength lies in the relationship between the curves which are drawn and the explanations that evaluators provide. |

useful images and stills from videos, and direct quotations from participants, can be in the preparation of summative evaluation reports as well.

For summative reporting, ISO 25062 (2006) provides a 'Common Industry Format (CIF) for Usability Test Reports' (copies can be purchased from <http://www.iso.org/>). The standard provides a report format that should help to improve the consistency with which usability evaluation is conducted and reported. The basic format of the report includes

- Description of the product
- Goals of the test
- Test participants
- Tasks that users were asked to perform
- Experimental design of the test



- Method by which the test was conducted
- Usability measures and data collection methods
- Numerical results

## EVALUATING FUTURE HCI

At one level, the evaluation of future HCI calls for the application of current evaluation techniques. When I say future HCI, I mean mobile and ubiquitous computing (including wearable computers), virtual and augmented reality and computer gaming. Thus, one can imagine the four reasons for evaluating HCI mentioned previously to be relevant to future HCI, that the four approaches to HCI will have a bearing on future HCI and that the five types of methods will be employed. However, there are other aspects of future HCI that call for rethinking of evaluation. In other words, it might not be entirely appropriate to take methods that have proven useful for evaluating desktop HCI and apply these to future HCI. Indeed, it is an open question as to whether the concept of 'usability' is entirely appropriate to future HCI; for example, how would one define the usability of a virtual reality (VR) experience? It seems to me that a problem in evaluating future HCI lies in defining adequate referent models. As Wilson and Nichols (2002) point out, 'There are only a limited number of ways in which we can assess people's performance generally and in Virtual Environments (VEs). We can measure the outcome of what they have done, we can observe them doing it, we can measure the effects on them of doing it or we can ask them about either the behaviour or its consequences' (p. 1032). At one level, this is simply because future HCI is attempting to develop approaches to interaction with technology for which there are no existing models. As mentioned earlier, one way to deal with this problem is to focus on activities that people are performing using a variety of products. However, this will only cope with part of the problem. For instance, the electronic tour-guide given earlier could be evaluated in comparison with other ways of performing activities, but this does not tell us whether any differences between the electronic tour-guide and the other products are due to the concept or to the realisation of the concept. In other words, if we find that the electronic tour-guide performs less well than speaking to someone, is this because the tour-guide lacks information, or because it lacks clear presentation of information, or because it lacks speedy access to the information, or because of some other reason (our evaluation would only point to all of these, not to specific reasons).

At another level, evaluation of future HCI means reconsidering and redefining context of use. Thus, in a virtual environment, one might need to consider technical factors such as rendering, frame rate and image quality, and subjective responses such as comfort, presence or credibility. In VR, there has been a great deal of research into evaluating the comfort (or more appropriately the discomfort) associated with the use of VR equipment and users' responses to the simulated environments, particularly in terms of nausea. Much of this work has resulted in user surveys, for example, into subjective responses to VR from the perspective of 'simulator sickness' or 'fidelity'. For example, Cobb et al. (1999) report a set of tools to evaluate the impact of VR on users termed VRISE (Virtual Reality Induced Symptoms and Effects). Parts of this toolset (e.g. a post-exposure symptoms checklist) are given to people following immersion in a virtual environment and are used to score the experience, particularly in terms of the potentially negative aspects such as nausea, disorientation and visual fatigue. A major focus of research in the VR community, certainly in the 1990s, was the definition and measurement of 'presence'. Again, much of this work has led to the development of self-report questionnaires (e.g. Slater et al., 1994; Witmer and Singer, 1998). The issue of whether presence (or any other phenomenon that can be associated with the experience of engaging with a VR) can be measured in a fashion that can use more objective measures continues to be a matter of some debate. Recent years have seen a growth in consideration of usability for games technologies; for example Pinelle et al. (2008) discuss heuristic evaluations which can be applied in the design and development of games based on a review of examples of gameplay that have been captured by users and posted on to gamespot. The tendency of gamers

to post either successful plays or ‘epic fails’ on to game sites provides a valuable repository for researchers, designers and evaluators.

In the field of mobile and wearable computers, much of the evaluation research has focused on comparing performance on a wearable computer with performance using other media. For example, a common observation is that people using the wearable computer tend to follow the instructions laid out on the display, whereas people using paper tend to adopt a more flexible approach (Siegel and Bauer, 1997; Baber et al., 1999). Of course, the notion that technology influences the ways in which people work is often taken as ‘common sense’ by ergonomists. However, the question of how and why such changes arise ought to have a far deeper impact on evaluation than is currently the case. In addition to evaluating performance, it is possible to consider a wearer’s response to the technology. For example, Knight and Baber (2005) report a comfort rating scale that can be used to gauge the subjective response to wearing and using such technology. As these technologies develop further and begin to find niche or mainstream applications, then the challenges of evaluation continue to develop. While the usability or UX of these new technologies can be considered using approaches explored in this chapter, they also raise questions about the integration of such technology into other activity. Not only does this relate to questions about attention-sharing and workload (Chapter 18) but also it relates to distraction and interruption for the user. Understanding the impacts of distraction or interruptions from these ‘always-on’ technology will not only help understand the impact of such technology on user activity, but also assist in the design of future technologies which are specifically designed to support ‘intelligent interruption’ (i.e. providing information to the user at points when they are likely to welcome the information or the interruption).

One of the current dilemmas in evaluating future HCI lies in establishing appropriate benchmarks for comparison. After all, the point of these technologies is to move beyond the conventional desktop-bound personal computers and to create new forms of interaction. However, the move to very different technologies makes it hard to establish a sensible basis for evaluation. Recall that it was proposed that evaluation requires a comparison with a referent model. So, what is the referent model for future HCI? As we have just seen, wearable computers often take paper as a referent model; the idea is that paper represents the ‘current’ way of working. Rather than solving the problem of a referent model, this, I think, only serves to bring the problem into stronger relief. Providing the two (or more) conditions of an experiment require similar actions and use similar technologies, then one can have some faith in the results. As soon as the actions or technologies begin to change, then care must be taken to ensure that these changes are catered for so as not to induce artefactual results. This is not to say that we cannot produce well-grounded evaluations of future HCI, but that care needs to be taken to ensure that we make use of a referent model, that is we do not simply ask people who have used our new ‘ghee-whizz’ device ‘do you like it?’ and expect to get sensible, useful and valid data.

## DISCUSSION

The key concept in HCI evaluation, ‘usability’, is seen as a multidimensional concept, and each dimension has different measurement requirements. Indeed, Frøkjær et al. (2000) demonstrate that effectiveness, efficiency and satisfaction need independent measurement in order to produce satisfactory results. The International Standards have been working towards such an approach, although they tend to give priority to a quantitative approach at the expense of other approaches. Having said this, the focus of UX offers an alternative approach to evaluation, which focuses on the qualitative, subjective experience of users. Rather than regarding these as distinct approaches, one ought to regard them as part of the continuum over which HCI can be evaluated. Whichever approach they take, evaluators should clearly state what they are evaluating, how they are evaluating, why they are evaluating things in this manner and what form of referent models is being used. What is always important is that evaluation involves some notion of ‘critique’ – the aim is not to ‘prove’ that the design is OK, but to ask how it could be improved. In other words, the rationale for undertaking

evaluation should not be simply conformance testing but should focus on improving the design or asking how the design fails to support specific tasks or to meet specific design targets.

## REFERENCES

- Andre, T.S., Williges, R.C. and Hartson, H.R. 1999. The effectiveness of usability evaluation methods: Determining the appropriate criteria. In: *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society, pp. 1090–1094.
- Baber, C., Arvanitis, T.N., Haniff, D.J. and Buckley, R. 1999. A wearable computer for paramedics: Studies in model-based, user-centred and industrial design. In: M.A. Sasse and C. Johnson (eds.), *Interact'99*. Amsterdam, the Netherlands: IOS Press, pp. 126–132.
- Baber, C. and Stanton, N. 2001. Analytical prototyping of personal technologies: Using predictions of time and error to evaluate user interfaces. In: M. Hirose (ed.), *Interact'01*. Amsterdam, the Netherlands: IOS Press, pp. 585–592.
- Bargas-Avila, J.A. and Hornbaek, K. 2011. Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. In: *CHI'11*. New York: ACM, pp. 2689–2698.
- Bevan, N. 2001. International standards for HCI and usability. *International Journal of Human-Computer Interaction*, 55, 533–552.
- Brooke, J. 1996. SUS: A quick and dirty usability scale. In: P.W. Jordan, B. Weerdmeester, B.A. Thomas and I.L. McLelland (eds.), *Usability Evaluation in Industry*. London, U.K.: Taylor & Francis Group, pp. 189–194.
- Card, S.K., Moran, T.P. and Newell, A. 1983. *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: LEA.
- Chin, J.P., Diehl, V.A. and Norman, K.L. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In: *CHI'88*. New York: ACM, pp. 213–218.
- Cobb, S.V.G., Nichols, S.C., Ramsey, A.D. and Wilson, J.R. 1999. Virtual reality induced symptoms and effects (VRISE). *Presence: Teleoperators and Virtual Environments*, 8, 169–186.
- Desmet, P.M.A., Porcelijn, R. and van Dijk, M.B. 2007. Emotional design; application of a research-based design approach. *Knowledge Technology Policy*, 20, 141–155.
- Dix, A., Finlay, J., Abowd, G. and Beale, R. 1993. *Human-Computer Interaction*. Hemel Hempstead, U.K.: Prentice-Hall.
- Frøkjær, E., Hertzum, M. and Hornbæk, K. 2000. Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 345–352.
- Good, M., Spine, T.M., Whiteside, J. and George, P. 1986. User-derived impact analysis as a tool for usability engineering. In: *CHI'86*. New York: ACM, pp. 241–246.
- Gray, W.D., Atwood, M.E., Fisher, C., Nielsen, J., Carroll, J.M. and Long, J. 1995. Discount or disservice? Discount usability analysis – evaluation at a bargain price or simply damaged merchandise? In: *CHI'95*. New York: ACM, pp. 176–177.
- Gray, W.D., John, B.E. and Atwood, M.E. 1993. Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world performance. *Human-Computer Interaction*, 8, 237–309.
- Gray, W.D. and Salzman, M.C. 1998. Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13, 203–261.
- Greenberg, S. and Buxton, B. 2008. Usability evaluation considered harmful (some of the time). In: *CHI'08*. New York: ACM, pp. 111–120.
- Hassenzahl, M. and Tractinsky, N. 2006. User experience – a research agenda. *Behavior and Information Technology*, 25, 91–97.
- Hertzum, M. and Jacobsen, N.E. 2001. The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, 13, 421–443.
- Holcomb, R. and Tharp, A.L. 1991. What users say about software usability. *International Journal of Human-Computer Interaction*, 3, 49–78.
- ISO 9241. 1998. Ergonomics of office work with VDTs-guidance on usability. Geneva, Switzerland: International Standards Office.
- ISO 13407. 1999. Human-centred design processes for interactive systems. Geneva, Switzerland: International Standards Office [superseded by ISO 9241-210 (2010)].
- ISO 9126. 2000. Software engineering – product quality. Geneva, Switzerland: International Standards Office.
- ISO 9216. 2001. Software engineering – product quality, [superseded by ISO 25010 (2011) Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – System and Software quality models]. Geneva: International Standards Organisation.

- ISO 25062. 2006. Software Engineering – Software product quality requirements and evaluation (SQuaRE) – common industry format (CIF) for usability test reports. Geneva: International Standards Organisation.
- ISO-9241-210. 2008. Ergonomics of human-system interaction. Geneva, Switzerland: International Standards Office.
- Jeffries, R., Miller, J.R., Wharton, C. and Uyeda, K.M. 1991. User interface evaluation in the real world: A comparison of four techniques. In: *CHI'91*. New York: ACM, pp. 119–124.
- Kim, J.H., Gunn, D.V., Schuh, E., Phillips, B., Pagulayan, R.J. and Wixon, D. 2008. Tracking real user experience (TRUE): A comprehensive instrumentation solution for complex systems. In: *CHI'08*. New York: ACM, pp. 443–452.
- Kirakowski, J. and Corbett, M. 1988. Measuring user satisfaction. In: D.M. Jones and R. Winder (eds.), *People and Computers IV*. Cambridge, New York: Cambridge University Press, pp. 329–430.
- Kirakowski, J. and Corbett, M. 1993. SUMI: The software usability measurement inventory. *British Journal of Educational Technology*, 24, 210–214.
- Kjeldskov, J. and Graham, C. 2003. A review of mobile HCI research methods. *Proceedings of the Fifth International Mobile HCI Conference*. Berlin, Germany: Springer, Lecture Notes in Computer Science.
- Kjeldso, J., Skov, M.B., Als, B.S. and Høegh, R.T. 2004. Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In: *Mobile HCI 2004*. Berlin, Germany: Springer, Lecture Notes in Computer Science, pp. 61–73.
- Knight, J.F. and Baber, C. 2005. A tool to assess the comfort of wearable computers. *Human Factors*, 47, 77–91.
- Kujala, S. and Väänänen-Vainio-Mattila, K. 2009. Value of information systems and products: Understanding the users' perspective and values. *Journal of Information Technology Theory and Application*, 9, 4.
- Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karanos, E. and Sinnelä, A. 2011. UX curve: A method for evaluating long-term user experience. *Interacting with Computers*, 23, 473–483.
- Macleod, M. and Rengger, R. 1993. *The Development of DRUM: A Software Tool for Video-Assisted Usability Evaluation*. Teddington, U.K.: National Physical Laboratory.
- Newman, W.M. and Lamming, M. 1995. *Interactive System Design*. Reading, MA: Addison-Wesley.
- Nielsen, J. 1993. *Usability Engineering*. London, U.K.: Academic Press.
- Nielsen, J. and Levy, J. 1994. Measuring usability: Preference vs. performance. *Communications of the ACM*, 37, 66–75.
- Norman, D.A. 1988. *The Design of Everyday Things*. New York: Basic Books.
- Olson, J.R. and Olson, G.M. 1990. The growth of cognitive modelling in human-computer interaction since GOMS. *Human-Computer Interaction*, 3, 309–350.
- Pinelle, D., Wong, N. and Stach, T. 2008. Heuristic evaluation for games: Usability principles for video game design. In: *CHI'08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 1453–1462.
- Ravden, S.J. and Johnson, G.I. 1989. *Evaluating Usability of Human-Computer Interfaces*. Chichester, U.K.: Ellis Horwood.
- Shackel, B. 1984. The concept of usability. In: J. Bennett, D. Case, J. Sandelin and M. Smith (eds.), *Visual Display Terminals: Usability Issues and Health Concerns*. Englewood Cliffs, NJ: Prentice-Hall, pp. 45–88.
- Siegel, J. and Bauer, M. 1997. On site maintenance using a wearable computer system. In: *CHI'97*. New York: ACM, pp. 119–120.
- Slater, M., Usoh, M. and Steed, A. 1994. Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 3, 130–144.
- Smith, S.L. and Mosier, J.N. 1986. Guidelines for designing user interface software. Report ESD-TR-86-278. Bedford, MA: The Mitre Corporation.
- Sonderregger, A., Zbinden, G., Uebelbacher, A. and Sauer, J. 2012. The influence of product aesthetics and usability over the course of time: A longitudinal field experiment. *Ergonomics*, 55, 713–730.
- Spool, J. and Schroeder, W. 2001. Testing web sites: Five users is nowhere near enough. In: *CHI'01*. New York: ACM.
- Springett, M. 2008. Assessing user experience within interaction: Experience as a qualitative state and experience as a causal event. In: E. L.-C. Law, N. Bevan, G. Christou, M. Springett and M. Lárusdóttir (eds.), *Meaningful Measures: Valid User Experience Measurement*. Toulouse, France: Institute of Research Informatics in Toulouse, pp. 86–90.
- Stanton, N.A. and Baber, C. 1992. Usability and EC 90/270. *Displays*, 13, 151–160.
- Stanton, N.A. and Young, M.S. 1999. *A Guide to Methodology in Ergonomics*. London, U.K.: Taylor & Francis Group.
- Suchman, L.A. 1988. *Plans and Situated Actions*. Cambridge, U.K.: Cambridge University Press.

- Vermeeren, A.P.O.S., Law, E.L.-C., Roto, V., Obirst, M., Hoonhout, J. and Väänänen-Vainio-Mattila, K. 2010. User experience evaluation methods: Current state and development needs. In: *NordiCHI2010*. New York: ACM, pp. 521–530.
- Whiteside, J., Bennett, J. and Holtzblatt, K. 1988. Usability engineering: Our experience and evolution. In: M. Helander (ed.), *Handbook of Human-Computer Interaction*. Amsterdam, the Netherlands: Elsevier, pp. 791–817.
- Wilson, J.R. and Nichols, S.C. 2002. Measurement in virtual environments: Another dimension to the objectivity/subjectivity debate. *Ergonomics*, 45, 1031–1036.
- Witmer, B.G. and Singer, M.J. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7, 225–240.
- Wright, P. and Blythe, M. 2007. User experience research as an interdisciplinary: Towards a UX manifesto. In: *British Computer Society HCI'07*. Cambridge University Press, U.K., pp. 65–07.
- Young, R.M. and Whittington, J. 1990. Using a knowledge analysis to predict conceptual errors in text-editor usage. In: *CHI'90*. New York: ACM, pp. 91–96.





---

# 14 Analysing and Designing Control Facilities

*Rebecca Charles, Sarah Sharples, Jane A. Rajan,  
John R. Wilson and John Wood*

## CONTENTS

|                                                                           |     |
|---------------------------------------------------------------------------|-----|
| Introduction.....                                                         | 384 |
| Human–Machine Interface.....                                              | 387 |
| Control Room Context.....                                                 | 388 |
| Process for Analysis and Design of Human–Machine Interfaces.....          | 390 |
| Initial Analysis .....                                                    | 391 |
| User Needs and Task Analysis .....                                        | 391 |
| Mental Model, Expertise and Behaviour Assessment.....                     | 393 |
| Environment and Organisational Influences, and Circumstances of Use ..... | 396 |
| Outline System Design .....                                               | 397 |
| Information Analysis.....                                                 | 397 |
| Operator Demands and Workload.....                                        | 398 |
| Control Dynamics Capability and Requirements.....                         | 399 |
| Modality and Interaction Preferences and Needs.....                       | 400 |
| Detailed Element Design .....                                             | 406 |
| Control and Display Specification and Integration .....                   | 406 |
| Display Formats .....                                                     | 406 |
| Mimic Displays .....                                                      | 407 |
| Sequence Displays.....                                                    | 407 |
| Deviation Displays .....                                                  | 407 |
| Alphanumeric Displays .....                                               | 407 |
| Alarms and Warnings .....                                                 | 408 |
| Graphical/Trend Displays.....                                             | 408 |
| Polar Coordinate/Object-Oriented Displays.....                            | 409 |
| Pictorial Display .....                                                   | 409 |
| Prototyping and Evaluation.....                                           | 410 |
| Integration and Evaluation .....                                          | 410 |
| Integration into Control Context .....                                    | 410 |
| Conclusion .....                                                          | 413 |
| References.....                                                           | 413 |

## INTRODUCTION

On 20 April 2010, the *Deepwater Horizon* mobile offshore drilling unit suffered an explosion and a subsequent fire that killed 11 workers and injured 16 more. The rig was drilling into a well, a process which involves creating a deep borehole and filling the casing with cement. This cement barrier failed, which allowed hydrocarbons and mud from the reservoir to escape through the drill pipe, causing the explosion.

Prior to the explosion, several tests were carried out which were interpreted by the crew as successful. One of these was a negative pressure test for which the crew used a process that was not the normally preferred method. Having interpreted this negative pressure test as successful, when the initial alert to the barrier failure was displayed (a high drill line pressure reading), the crew assumed that this high reading was nothing to worry about and carried on with normal activities. Whilst carrying out these activities, the crew failed to notice the drill pipe pressure increase even further having not been informed of the correct procedures to follow in terms of monitoring during these activities. At this point, operators noticed mud flowing onto the floor of the rig and attempted to divert the hydrocarbons to the mud gas separator. Unfortunately, the mud gas separator was designed for use in low-pressure situations, and this approach was therefore unsuccessful due to the high pressure. At this stage in the incident, a decision to pump the mud and hydrocarbons overboard could have been taken, but this was not done, and the gases caused an explosion. The accident report concluded that the incident was not down to a single cause and was in fact the culmination of complex interactions of human judgements, team communication, mechanical failures, engineering design and operational implementation (Hubbard and David, 2010).

On Sunday, 10 December 2005, a tank at Buncefield oil storage depot in Hertfordshire exploded and resulted in a fire that burned for 5 days. The tank in question was tank 912, with a capacity of 6 million litres, and on the night of the incident was filling with petrol. The filling could be monitored by employees via a gauge, the output of which was visible on a computer screen, and was backed up by an independent high-level switch (IHLS) that would apparently prevent the tank from being overfilled. In the early hours of the morning, the gauge failed, leading the display to flat line, and the tank continued to fill. No alarms were sounded due to the failed gauge, so the supervisor was unaware of the situation. When the tank reached a certain level, the IHLS should have been tripped to close the valves preventing the tank from overfilling; it did not. Approximately 2½ h after the gauge display failed, the tank exceeded its capacity and petrol started to spill out. A large vapour cloud formed above the tank and was noticed by members of the public. On-site employees were notified, and the fire alarm button was pressed. A spark from the firewater pump exploded the vapour cloud and ignited the tank along with 250,000 L of escaped petrol. The fire engulfed over 20 tanks on the site, and chemicals and fuel penetrated the soil beneath a number of drains, and soakaways near the site. Remarkably, partly due to the time of the day on a Sunday at which the incident occurred, no one was killed, but over 40 people (mainly emergency services attending) presented symptoms consistent with smoke and fume inhalation.

The failure of the gauge system was quoted as an immediate cause of the incident (COMAH, 2011). The gauge display was on a screen that was used by several more tanks, and only one tank could be viewed on the screen at a time. On the night of the incident, the gauge display for tank 912 was behind other tank displays. The operators relied heavily on alarms alerting them to any abnormality. The flatlining that occurred on that night had in fact occurred 14 times as a *false alarm* in the 3 months leading up to the incident. This became so commonplace that faults were often not logged. In addition, the IHLS was incorrectly installed and maintained, and staff were unaware of the full operating procedure. As well as the immediate causes of the incident, the wider managerial failures surrounding the incident also played a major part (COMAH, 2011) and highlight the complexity of such incidents.

In these cases and countless other examples, we find design of the displays and controls making up the human-machine interface (HMI), even at a very basic level of size and positioning, in addition to any issues involving information support for problem solving and reasoning. Many displays and controls that we still find within our current-day workplaces are particularly inadequate in the way they can lead operators to make errors or lead to the reduce violation of safety procedures. Poor interface design and implementation can manifest itself in physical problems (e.g. the display cannot be seen), cognitive problems (e.g. a coding is not understandable) or organisational problems (e.g. two team members have a different understanding about what is really happening).

At the heart of human-machine systems is the exchange of information between people and the system through displays and controls. Indeed, the early years of ergonomics/human factors (E/HF) are often characterised by the expression *knobs and dials ergonomics*. The interaction of person and process is at the centre of a total systems view, and displays and controls provide the media for that interaction and for carrying out tasks. According to the type of system and the view of the analyst, the HMI might physically or conceptually lay between the person and the displays and controls, or between the controls/displays and the machine, or might comprise the controls and displays themselves. In the first of these views, the interface is seen as a conceptual entity, in the third as a physical entity and in the second as either or both. In virtual reality and virtual environments, the interface can be seen as being ubiquitous, comprising the interaction experience itself.

In this chapter, we consider the challenges and processes associated with analysing and designing control interfaces. How we define the interface depends on where we draw the system boundary. For instance, a book can be seen as a *read only* display system, or we can see reading a book or a set of instructions as a multiple-person system with a temporal dimension: one person uses controls – such as a pen or computer – to produce a display of information which is seen later, on one or several occasions, by themselves or by other people. Controls and displays that make up the HMI take a wide variety of forms, and this in large part is the reason for the awkward way in which authors often refer to controls and displays of a *process* or *system* in order to avoid having to refer explicitly to tools, equipment, machines or products.

It is unfortunate that there is no obvious word which suggests all types of method for communicating between the operator and the plant, or one operator and another, because many human factors recommendations are the same whatever the technology used. For example, most recommendations about the shape of numbers which can be read most quickly and accurately apply whether those numbers appear on conventional instruments, VDUs, controls, or printed reference materials. Questions about how easy controls are to use, apply whether the operators are indicating what action they want to make by turning a knob, pressing a switch, using a list of alternatives on a VDU, or making a telephone call. In the following... the word '*display*' refers to all methods of giving information to the operator, the word '*control*' applies to all means by which the operator gives instructions.

(Ball, 1991)

Amongst the variety of interface forms, we might find the following:

- Industrial machines which have hand or foot controls (pedals, levers, buttons) and displays of numerical quantities or qualities (states).
- Computer systems which can utilise input devices operated by hand, foot or even head (for the disabled) or where input may be achieved *directly* through speech, gestures, eye movements, etc., but where information presented to the user is largely by visual display with some auditory feedback; personal devices, with biometric input and output, extend on these channels.
- Simulation and virtual environment systems which aim to give the user some feeling of being in the world which is modelled, where they may have no control, or control via

a veridical interface (driving or aviation simulator), or control via novel input devices (data-glove or spaceball in virtual reality, for instance).

- Augmented reality in which real objects are added to virtual ones which could enable prediction in certain situations.
- Gesture systems that can allow the manipulation of the system through gesturing where hand signals and movements can control the system.
- Display-only systems, usually with visually displayed information (e.g. signs, books) but which also can be auditory (warnings) or tactile (e.g. Braille).
- Products where the *control* is built into or is part of the product, and information is displayed back in part through visual confirmation of the control action but also tactilely, kinesthetically or proprioceptively (e.g. toothbrush, razor, screwdriver, hammer, manual gear change, game controls).
- Touch screens that allow the operator to monitor, alter and update the display by directly manipulating the interface and are provided with visual and often tactile (vibration) feedback.

Transportation where, although traditional-type controls and displays exist, the greatest part of information is received by directly seeing the course or track (road or river) and registering the vehicle's state along it.

As technology develops, these systems become increasingly multimodal, and the operator is able to select the preferred mode of interaction, for example, in groupware conferencing systems offering video, whiteboard, messaging, group calendars and tools and voice communications.

We will use the convention in this chapter of defining the interface as being the medium through which the two-way exchange of information takes place, between people and computers and between people and other people via computers and telecommunications. The chapter will not address all types of interface but will focus on the control and display issues generally associated with control panels and control rooms in process and transport control. Evaluation of human-computer interfaces is dealt with in detail in Chapter 13. Here we will deal with the sorts of control and display issues which include conventional instruments – analogue (quantitative and qualitative) and digital, annunciators, alarms, computer-generated displays, information from written or verbal sources, information from the *state* of a machine, process or environment, and all forms of controlling a process or system. Within our review, we will introduce the need for a considered approach, both when analysing existing HMIs and when designing and evaluating new ones, and present some of the theoretical perspectives that it is important to consider when undertaking such design. Following a brief overview of the HMI and control room issues, we discuss at length a set of activities that must be carried out in any analysis and design of interfaces for control rooms. Then, some detailed discussion of interface specification, especially for displays, covers choice of display mode, display formats, coding and navigation and structuring of displays, with consideration of the increasingly multimodal nature of such displays and systems. Finally, although we cannot cover all issues of evaluation, we provide a basic control room interface evaluation checklist. A more comprehensive list can be found in ISO 11064-7. As such the operator is an integral part of the system and thus for the system to be both safe and perform effectively, this human component must be designed as part of the system.

Many design engineers accept the arguments (of human based design) but when safety by design becomes difficult they relapse into saying 'we shall have to rely on the operator'. This is the essential issue in ergonomics. If reliance on the operator is necessary in order to assure that the plant design is safe and reliable one has to first understand how the human behaviours of that operator that provides the critical link in plant functioning (which in such cases is often safety critical).

(Kletz, 1991)

The scope of HMI issues specific to control rooms is addressed in ISO standard 11064 (in particular 11064-5) Ergonomic Design of Control Centres: this set of standards provides information on principles pertinent to the ergonomics design of control rooms. Parts 1 and 2 cover the principles of the design and arrangement of control centres and suites; part 3 focuses on control room layout; part 4 looks at workstations; part 5 covers displays and controls; part 6 considers environmental issues and requirements; and part 7 considers principles for the evaluation of control centres. Other relevant ISO standards include the following:

- ISO 9241 Ergonomics of Human-System Interaction
- ISO 9355 parts 1 and 2, ergonomic requirements for the design of display and control actuators

Chapter 37 presents a more in-depth discussion of standards of relevance to E/HF.

To consider control–display interfaces and control room ergonomics, other useful sources of information, some a little dated but still with useful information and guidance, include (Ball, 1991; Gilmore, 1989; Goodstein et al., 1988; International Instrument Users Association, 1998; Ivergard and Hunt, 2008; Kinkade and Anderson, 1984; Kirwan, 1994; Moray, 1997; Noyes and Bransby, 2001; NUREG-0700, 1996; NUREG-5908, 1994; Sheridan, 2002; Stanton, 1994; U.S. Department of Defence, 1999; Wood, 1994, 2004).

## HUMAN–MACHINE INTERFACE

A classic view of the interface has been to understand the operator (person controlling the interface) as a passive and limited capacity processor of information. In this view, the operator and *machine* are in a closed loop (although comprising an open system), connected by distinct displays and controls. *Machine information* is converted into *operator information* via displays, and controls act as transducers to allow the operator to change a system state. Feedback to the operator comes via the displays and via interaction with the controls.

New technologies and system designs require a change to this view. First, control systems are increasing in their *intelligence* – they not only act as a communication conduit between the operator and the system under control but also interpret information and make decisions and suggestions. This means that the operator not only must have a full understanding of the system under control but should also understand the way in which the information has been interpreted or analysed before it is presented. An example of this can be found in the rail signalling context. Interviews with rail signallers (Balfe et al., 2012) revealed that whilst signallers were effective and safe in their use of an automated system to support allocation of trains to routes, they did not fully understand the way in which the technology made the decisions about which trains to route in which order; as a consequence, this understanding affected the way in which they chose to interact with the system. A second evolution in control systems is the lack of distinction between displays and controls. The increasing ubiquity of touch screen interfaces (Sharples, 2012) has led to much less distinction between the *display* and *control* elements of a system – in many cases, the two are now the same visual entity. Finally, control systems continue to increase in their complexity, introducing increasing numbers of displays, and also linking handheld, desktop and wall-based displays and a single connected system. These are just some examples of changes that necessitate a change in emphasis and approach for control room design.

Skill requirements in perceptual judgement, decision making, problem solving and diagnosis have, in the past, led to more sophisticated models of an operator. From the work of authors in the collections by Edwards and Lees (1974), we can see an expansion of the original simple operator–process loop. The way people interact with systems is modelled as including attributes of the operator such as their mental model and experience, and includes representation of their interaction

through formal and informal procedures. The model is frequently conceived of as the *human as supervisory controller* (Sheridan, 1987). In this view, computer systems mediate between the operators plus their displays and controls on the one hand and the task or process and its sensors and actuators on the other. Sheridan identifies ten cause–effect loops in supervisory control and defines possible supervisory roles for the operator as planning, teaching, monitoring, intervening and learning, so we can see the need for a structured comprehensive approach to the design of display–control interfaces. If system control can occur in so many ways, and the needs and roles of operators can be so varied, then interface analysis and design must be based on an appreciation of much more than checkpoints for selecting individual instruments. In addition, the interaction with any given element of the interface may be fluid in its nature, and interfaces may adapt or be adapted to suit the needs of each individual operator.

Hollnagel and Woods emphasise the need to look beyond this traditional *input–output* approach by considering people, technologies and artefacts working together as a joint cognitive system (Hollnagel and Woods, 2005). This is also reflected in the concept of distributed cognition (Hutchins, 1995) which looks at the transmission of information in a distributed context and has been used as a paradigm to examine contexts including cockpits (Hutchins and Klausen, 1996) and ambulance dispatch (Blandford et al., 2002).

## CONTROL ROOM CONTEXT

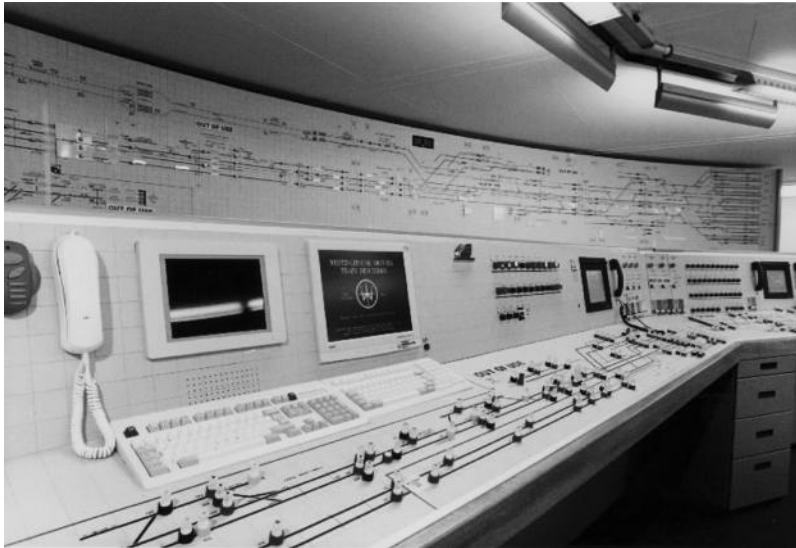
With the increasing use of computers to control processes and systems, the tendency for the control of such systems to be centralised has continued, usually from one primary location to a central control room. Control rooms are found in many different domains – transport systems (e.g. air traffic control, railways, metros), emergency services (police, fire, ambulance), industrial processes (steel, chemical, food) and power plant (nuclear, electricity), security applications (banks, prisons, public or private buildings). Importantly, although the term used generally and in this chapter is *control room*, these may in fact be a series or suite of rooms.

‘A control room or control centre is the place where one or more people, sitting at control desks, conduct control activities. A control suite... is a group of co-located, functionally related rooms, including a control room... [and] control complex is... a series of functionally related rooms which are on different sites...’ (Wood, 1994).

The complexity of technological systems has continued to grow so the demands for the safety and efficiency of these systems have grown apace. Increasingly, the overall reliability of such systems as limited by the reliability of the hardware components has increased (although perhaps not as much as engineers might like to think!), so human operators can become the most unreliable components of a system, unless appropriate ergonomics input is made. It is imperative that the interfaces which provide the means of interaction between the system and the people who operate them allow the operation of the system to be as efficient and error free as possible. As systems become more complex and increasingly automated, the requirements placed on the system operators are evolving. Each group of operators, or in some cases each operator, may have choices about the configuration of the HMI they interact with the system through. The allocation of function and level of automation may not be fixed but flexible. This flexibility may in its turn be automated (e.g. if an emergency scenario occurs, the operator may not be allowed to have any direct control over the system for a defined period of time, to allow him or her to gain a more complete understanding of the current situation). Alternatively, the operator may select the use of automation, for example switching autopilot on, for navigation of an aeroplane or a ship. This may support periods of high workload or allow other tasks to be given more attention. This chapter addresses the issues to be considered in the design of interfaces to support the operation of such complex systems.

In line with the increase in system complexity has come an increase in the amount of information available to the operator to support system operation and the capability to display this information in a variety of formats. Control room interfaces are increasingly based around several visual





**FIGURE 14.1** Example of hard-wired panel display in rail.

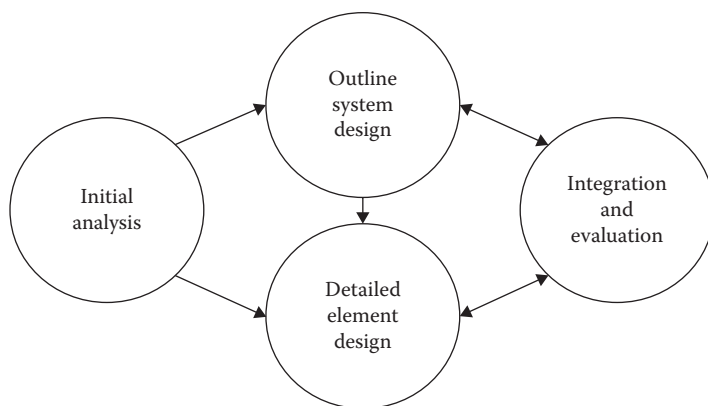
display unit (VDU) displays – rather than wall or console mounted panel displays which use dials and similar display types to present the information to the operator (Figure 14.1); although physical control panels are still found and implemented in some contexts. This has had a significant impact on the design of control room interfaces. No longer is all the information concerning system parameters always available to operators at the same time. Visual displays give far greater flexibility in the presentation of the information: the same data can be presented in a variety of different formats, but there is a limited amount of information that can be presented on one display screen. To overcome this, one operator may have several displays from which the system can be operated. ‘How many screens do the operators need?’ is one of the questions most often asked by control room designers (see also Parts 4 and 5 of ISO 11064). It must be remembered that, the operator has to deal with the complexity of both the plant/process and also the control and surveillance systems. For example, in a chemical manufacturing plant, an operator may be required to control and monitor the process parameters such as flow rate (as with the Buncefield example earlier in the chapter), pressure and temperature on VDU-based process displays whilst also monitoring CCTV pictures of the plant that show visually the equipment and any personnel on the plant. One solution to this is to introduce analysis tools and algorithms to interpret information before it is presented to the operator; this human–technology hybrid approach is increasingly termed *intelligent infrastructure* (Dadashi et al., 2011). An intelligent infrastructure system will combine information from a range of sources, interpret and present it to the operator in a form that should directly affect decision making. An example of intelligent infrastructure is found in the monitoring of large distributed systems, such as the electrical system that supports much of the rail network in the United Kingdom. An individual control room has responsibility for a wide geographical area and receives alarms and alerts to highlight where problems or faults might be occurring. These alarms and alerts are drawn from increasingly sophisticated sensors that are placed on the network itself. Information from different sensors can then be presented in a combined form to the operator in terms of the likelihood of the presence of a fault, and advice regarding the need for inspection or intervention. This approach of combining and interpreting information potentially allows more information to be displayed in a smaller display, but the consequence of potentially losing the underlying and detailed information on operator performance needs to be carefully considered.

Overall, control rooms are increasingly collaborative, encompass a range of technologies and integrate a vast range of data sources. This complexity may be summarised as *multi-person*,

*multi-task, multi-technology and multi-artefact*. The downside of losing the large displays seen in many control rooms in the past is the reduction in shared information. Large displays allow someone to enter the control room and identify at a glance the state of the system and provide a regular pulse on activity. They also provide a main focus for visitors. These easily accessible displays were usually the only copy and would be viewed by all operators. One move, particularly in rail, is to remove the large mimics of track diagrams, in favour of VDU systems. This has the benefit of being able to control more railway from one location, and for an individual operator to control more railway, especially when combined with automation. Again as previously mentioned, this minimises the visibility for anyone entering the control room and could potentially impact in emergency situations. However, this is mitigated by supervisors having access to the same information in VDU form, but the amount of information able to be displayed is limited by the number of displays. A benefit of removing information from large mimics and moving it onto VDU systems are that the information can be accessed potentially from elsewhere and that information can be easily updated. For example, real-time signalling systems may be able to transmit information straight from the signalling system to the customers at the stations.

## PROCESS FOR ANALYSIS AND DESIGN OF HUMAN–MACHINE INTERFACES

We make no attempt to define a full process for control–display system design or control room design, which would have integral staged evaluations and iterations embedded within it. Nonetheless, the chapter is built around a set of activities which, taken in order, outline an analysis and design process. It is suggested that the procedures and activities shown in Table 14.1 comprise a logically ordered listing of what must be done to develop control/display interfaces for control rooms and human–machine systems generally, which has been used in practical control room human factors (although not all stages are necessarily included in each study or development). Any ergonomics approach to design will follow the same basic steps; for instance, current discussions about new versions and parts of ISO 11064 include consideration of a flow chart to inform and guide on the overall process, which matches very well the one we use here (see Figure 14.2). It is important to note that the selection of specific interface types and their development or their specification do not take place until very late in the process. The implication is that development cannot merely be built around design guidelines and equipment selection rules; before any such guidance can be used, the designer must gain a thorough understanding of user, task, system and environment requirements in a detailed analysis. Visiting equivalent environments may also be of benefit at this stage to gain a wider understanding of control room functions and layouts.



**FIGURE 14.2** Flowchart to guide the design process.

**TABLE 14.1**  
**Key Design Stages**

| Design Stages              | Design Substages                                                                                                                                                                                 |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Initial analysis           | 1. User needs and task analysis<br>2. Mental model, expertise and behaviour assessment<br>3. Environmental and organisational influences, circumstances of use and consideration of future needs |
| Outline system design      | 4. Information analysis<br>5. Operator demands and workload<br>6. Control dynamics capability and requirements<br>7. Modality and interaction preferences and needs                              |
| Detailed element design    | 8. Control and display specification and integration<br>9. Prototyping and evaluation                                                                                                            |
| Integration and evaluation | 10. Integration into control context                                                                                                                                                             |

The remainder of this chapter takes up discussion of the activities defined in Table 14.1. First we will examine the initial analysis activities – establishing user needs and current expertise, and understanding the organisational and environmental constraints that will be present in the control room setting. Second, there is a consideration of the different aspects of systems that need to be specified at the outline, earlier design stage and tools that can support this process. Third, some of the available guidance on interface and interaction specification is reviewed, concentrating upon display systems with particular application in control rooms. Finally, approaches to integrating and evaluating new systems in the control room context are discussed. As with other chapters in this book, this chapter is not intended to be a comprehensive *how to* instructive text, but to highlight the different approaches and considerations that should be addressed at different stages in the control room design process.

## INITIAL ANALYSIS

### USER NEEDS AND TASK ANALYSIS

As with all ergonomics analysis and design, we need first to establish the requirements for the system, to identify and describe tasks and users, which will involve carrying out a task analysis (see Chapter 6 in addition to the current section). It also involves making some first-level decisions on the functions that must be performed, and these are described at a gross level as activities that are needed to meet the system objectives (and usually collections of tasks). We need to consider the balance of responsibilities for functions between people and computers or other equipment (allocation of function) and between different people (division of function), including potential needs for automation. The outcomes and decisions made on the basis of such analysis can be returned to, reviewed and revised, in each subsequent stage, on the basis of new or amended information, ideas and opinions. Moreover, the task and user analysis themselves are not once-and-for-all exercises; they will be revised as appropriate in later stages on the basis of new information and decisions. However, it should be noted that task analysis should also be focused as there is a considerable danger of *overdoing* analysis and producing substantial amounts of information, which may not have direct application in the design process. Therefore, it is vital at the outset to have a clear, narrow focus about the key information requirements and what form the deliverable will take.

Where an existing interface is being assessed and analysed, many of the methods, measures and techniques described in this book will be appropriate, particularly verbal protocol analysis

(Chapter 7) since the work carried out from using information interfaces is often not directly observable. At this time, it is important to identify any likely significant constraints on the interface from the potential users or their tasks. Factors such as user experience, training and support must be predicted or assessed. Certain user attributes may determine a major decision at the outset. Disabled users are an obvious case, but other needs may be to design for a variety of cultures and languages, or to understand that military and civilian operators may well behave very differently from each other in circumstances where a technology is being transferred from defence to industrial application.

Often, there will be one operator or a team of operators who are highly skilled and knowledgeable about the system, but there may also be a variety of other users of the system, carrying out a wide range of activities, who need to be considered in the design of the interface. Such users include those responsible for maintenance, managers or supervisors who require information on system performance, and systems engineers who may require information on the way in which the control system itself is performing. For supervisors, an early decision will be required on the degree to which they are monitoring systems or operators, which will determine whether emphasis should be on equipment position for access, or on workstation configuration for communication. We must appreciate the different levels of operator knowledge involved in working with any particular system, knowledge we may wish to provide, support, enhance or merely to be aware of in designing interfaces. Wirstad (1988) provides a listing of such potential knowledge in process control, including categories of plant layout:

- Function, construction and capacity of components
- Manoeuvring (location and operation of controls and displays)
- System function, construction and flows
- Process theory and practice
- Identification of disturbances and consequence prediction
- Measures to take for disturbances
- Procedures for serious incidents
- Organisation structure
- Administrative arrangements
- Safety regulations
- Supervision

As well as knowledge required of operators, we must also be aware of all the tasks that must be supported by the interfaces (e.g. Table 14.2).

For interface design, there are a variety of analysis methods that can be applied to analyse the information requirements of the users; these need to show who will use the information, when it will be used, what it will be used for and how it will be used. Wood (2004) identifies seven types of

---

**TABLE 14.2**  
**Tasks to Be Supported by the Interface**

|                           |                                                                                                                                                                |
|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Procedural                | Tasks that involve following a predetermined sequence of events.                                                                                               |
| Sensory motor             | The physical manipulation or use of input devices and haptic interfaces.                                                                                       |
| Communication             | The transmission of information between operators without the information being translated into another format (i.e. verbal communication or logs).            |
| Monitoring and prediction | The surveillance of the system to identify and anticipate any change in system status.                                                                         |
| Fault detection           | The identification of an abnormal or unexpected system status.                                                                                                 |
| Decision making           | The selection between alternative options/actions.                                                                                                             |
| Problem solving           | The process of resolving uncertainty about system states. A particular kind of problem solving that is especially relevant to this context is fault diagnosis. |

---

analysis that will be completed once the brief for the design project has been confirmed. These types of analysis are task analysis, functional analysis (applied to normal, degraded and maintenance contexts), operational simulations, hazard and risk analysis, systems interface analysis, workload analysis and human error analysis (Wood, 2004). Not all of these analyses will always be performed for all designs or redesigns, but it is important that the E/HF specialist is aware of the different elements of the current and prospective task that needs to be captured. Of course there is an interaction between these elements; the design of interfaces is dependent on the information available, which in turn is dependent, for example, on the system hardware and software. Whilst technology may mean that there is a large pool of potential data for display, it is often the case that the display format options are limited by factors including the underlying data structures within the system, the compatibilities of data, the space available within the control context and the cognitive capacity of the operator. As well as system feedback, consideration should be given to operator feedback. Supervisors may use subtle cues such as posture of operators, and changes in speech patterns to identify stress or high workload.

### MENTAL MODEL, EXPERTISE AND BEHAVIOUR ASSESSMENT

Once the initial analysis of the task and its elements has been completed, the specific characteristics of the operators need to be considered. This can be challenging, as operators can vary in their task completion strategies and behaviour; they develop experience over months and years in the same job and completing the same task, and increasingly, different groups of people, who may vary according to age or culture, may bring different past experiences, skills and expectations to a task. These need to be captured and taken into account in control system design, so that if for example, a new system requires new skills, appropriate training or design features are introduced.

The first aspect to be considered is the *mental model* that an operator brings to and develops whilst using a system. Imagine an operator performing remote control task who must carry out a job from a vantage point not used before, trying to get orientated and guide the system through critical operations in the most reliable, safe, effective manner; or the process worker who must try to understand why problems are occurring, and what combination of the many input parameters available would solve the problem, and must do this from process displays as well as by directly viewing the operations; or a maintenance engineer on a continuous process plant trying to relate information from system diagnostics to that on the control panel and to that on a giant mimic display. In all these cases, it is reasonable to think of the people involved as having or forming a mental model of the system. This mental model (or conceptual model) may be accurate or inaccurate, usable or worthless, but nonetheless the notion of mental models is attractive and useful for ergonomics.

In fact, in many cases, the worker will construct and use several mental models: perhaps a symbolic one of the interaction of the variables in the systems – electrical, or chemical, or mechanical; a pictorial one of the form of the system being worked with; and a model of rules governing the operation (correct or faulty) of the system. These models can be relatively concrete or abstract. The representations may be formed from the system itself, from operating, emergency or maintenance procedures, from instructions and training, or from other systems worked on in the past.

As regards methods to identify and represent mental models, much debate has focused on the degree of formality needed. Within cognitive psychology, there is the understanding that the notion has utility only if mental models can be described in computational form (e.g. Johnson-Laird, 1983). On the other hand, within human factors, we are usually willing to postpone questions of how people represent and use knowledge in favour of understanding what knowledge is represented and how it is used to make inferences in specific domains (Payne, 1988). Tools such as eye tracking are increasingly used to track visual attention, and it can be possible to ask operators to use a version of retrospective verbal protocol, to view and describe not only their performance with a system but also their eye movements whilst completing that task; this approach is one that can support the elicitation of mental models or their contributing factors. The majority of human factors literature,

and especially in human–machine systems, appears to refer to conceptual (or non-computational) mental models, that constitute the topographical, structural and functional understanding of a physical system and to describe, explain, understand and predicts system behaviour. It is the easiest to conceive such a model as comprising a *system simulation* that can be reconstituted and run in order to derive or confirm understanding.

There is some agreement about mental models that they:

- Are internal representations of objects, actions, situations, people, etc.
- Build on experience and observation, of the world and of any particular system
- Are simulations run to produce qualitative inferences
- Constitute topography, structure, function or operation of the system
- May contain spatial, causal, contingency relations
- Allow people to describe, predict and explain behaviour
- Underpin people's understanding and behaviour
- Are instantiated each time they are required, and are parsimonious, and therefore are incomplete, unstable and often multiple.

The last point, made by Norman (1983) and others, has high relevance for control/display interface design. For example, the models instantiated by, say, a central control room operator to assist in fault diagnosis at a flexible manufacturing cell will vary each time in type and content. If the problem is product quality related, then the cell may be conceptualised in functional form, modelled in terms of the series of transformation processes and the tooling needed to do this. Alternatively, if the problem is to do with hold-ups in components delivery, then the cell may be conceptualised physically and spatially in terms of element flows and bottlenecks. How operators arrange their screens may also be an indication of their mental models; different operators often have a preferred arrangement across their screens throughout their shift which may indicate differences in mental models. In many cases, more than one mental model may be formed, which will nonetheless overlap in their content and how they are employed, and which will have gaps or *inaccuracies* according to operator experience and training.

There are many serious questions about the notion of mental models, though, and little agreement on definition, identification, representation and utilisation or even adequacy in methodology for their identification (Rutherford and Wilson, 1991). But *if* we can predict or understand even in some fashion what mental models a new operator or user might hold about a system and its relevant domain, and what model they might build through subsequent interaction with the system, then we can improve interface design, training, operating procedures and so on. By understanding the potential users' mental models, and by adapting their own conceptual model accordingly, designers might develop a *system image* that better matches, sustains and helps develop an appropriate user mental model (Wilson and Rutherford, 1989). We must also acknowledge the varying impact of different levels of *expertise* on requirements for control room design.

The notion of behaviour, in terms of how tasks are completed, and the cognitive resources and activities involved, extends from the notions of mental models and expertise. Rasmussen and his associates (e.g. Goodstein et al., 1988; Rasmussen, 1986; Vicente 1999) distinguish between Skill, Rule and Knowledge based (S–R–K) behaviour. There is also an intimate relationship between expected task behaviours and the interactions completed to support them; in a basic description of how operators perceive information within the S–R–K model, Rasmussen uses a simple display/control system as illustration (Rasmussen, 1986, p. 107).

Skill-based behaviour is what is shown in tasks such as *in-the-loop* controlling or steering, adjusting and calibrating instrument settings or assembly tasks. The behaviour is akin to sensory-motor performance – where there is a fairly direct connection between sensory input and motor output with little mediation in cognition. In essence, skill-based behaviour is shown in familiar situations and is where the *operator* recognises a *signal* from the system and understands that this requires a



normal routine, then executes a well-learned skilled act more or less automatically. Generally, this type of work is undertaken with simple feedback control (comparison of the actual and intended states) that provides error information and thus defines the motor output response. However, there are suggestions that skill-based performance may also be based on feed-forward control and knowledge of the environment, for instance, when riding a bicycle.

Rule-based behaviour is said to require a more conscious effort than skill-based behaviour, and involves following a set of stored rules – in the mind or written down. Here, performance is goal oriented and takes place in familiar, routine and non-routine situations; the operator perceives a *sign* indicating environmental state(s), and then he or she uses learned rules and procedures once they have recognised certain cues. Control is fed forward through the stored rules. These rules will be derived empirically through experience or communicated to the operator by colleagues, institutions or training.

The boundary between skill- and rule-based behaviour is indistinct and may depend in the same situation on training or attention levels. A person working in skill-based mode may not be able to describe how they work, because the behaviour required may have become so automatic to them. The result of working in rule-based behaviour is frequently to coordinate and control a sequence of skill-based acts.

At the highest level is knowledge-based behaviour. Here the operator is in an unfamiliar situation with no, few or partial rules available from past experience. A goal is formulated by the operator, based on their perception of the state of the world and on some overall or global aims. Perceiving these symbols allows them to develop a plan, using knowledge, reasoning and experience, and the goal they themselves have formulated helps them work to that plan. The plan itself may be selected and tested through a process of conceptual or physical trial and error.

Rasmussen's model has been criticised, and subsequent authors have not always distinguished what they mean by *skill*, *rules* and *knowledge* clearly and consistently. On the other hand, a large number of important human factors studies have made use of the model and approach. Vicente (1999) gives a very useful account of its value in making cognitive task design decisions.

Within consideration of control–display interfaces, the least that might be said is that any displays must, in general, support all kinds of behaviour. For instance, the same information display might be needed to guide normal operations, routine maintenance and fault diagnosis in abnormal pressurised situations, and we have already seen that what starts out as a knowledge-based task may subsequently decompose into a rule- and then a skill-based task. Key elements in design to support skill-based behaviour include appropriate cues to prompt required skill-based activities (e.g. the presence of a clearly recognised green traffic light to indicate that a car driver should execute the skill-based behaviour of pressing the accelerator to move the car forwards) and avoidance of unnecessary distraction or interruption that might disrupt execution of skill-based behaviours. Rule-based behaviours may use both knowledge in the world (e.g. instructions or labels) and knowledge in the head (e.g. memorised sequences of operations); therefore, control interface design could consider this not only in providing quick to access and compatibly presented knowledge in the world but also to acknowledge that the driver is indeed executing rule-based behaviour, so presenting such information in a *rule-based* form (e.g. a table showing the types of alarm in the left-hand column, and the required response in the right-hand column). Knowledge-based behaviour relies on an individual appropriately assimilating information from the environment and using this information in combination with past experience to make decisions or select strategies. Awareness of the impact of decision-making biases (Kahneman and Tversky, 1984) is important in this case as the way in which information is presented will influence the information upon which an individual's attention is focused.

Ideally, not only should a display support all three types of behaviour, but it should also allow an operator to move *up and down* between levels, in terms of information detail and degree of abstraction. Furthermore, in many circumstances, it will be valuable for the system itself to support the operator by providing guidance on appropriate behaviour for different situations.

## ENVIRONMENT AND ORGANISATIONAL INFLUENCES, AND CIRCUMSTANCES OF USE

Although the person, process, controls and displays (which could be considered to be the *joint cognitive system* as discussed earlier) constitute to an extent a closed-loop sub-system at the heart of human-machine systems, this sub-system is itself open to the environment and context. Interaction will occur with all physical, psychosocial and organisational environment factors (see Chapter 1).

Wood (2004) identifies principles that should inform both the layout of a control room and the environmental context. His guidelines in these areas, with minor modifications, are presented in Table 14.3.

**TABLE 14.3**

### **Modified Guidelines from Wood (2004): Principles for Control Room Layout and Environmental Context**

#### *Control room layout*

- Room layouts should be based on a task analysis and an understanding of the worker population – including workers with disabilities
- Where control rooms house a number of operators, the layout should facilitate team working opportunities and social interaction
- Control room layouts should reflect the allocation of responsibilities and the requirements for supervision (e.g. appropriate grouping of team members)
- Operational links, including sightlines or direct speech communication, should be optimised when developing control room layouts
- Control rooms that exhibit either overcrowding of work positions or widely dispersed workstations are not recommended. Layouts should not only allow, wherever practical, direct verbal communications between control room operators but also avoid excessively short separation between adjacent operators
- Control rooms with similar functions, and in the same plant or facility, should adopt the same ergonomics principles of room layout to facilitate decision making and transfer of training
- Where physically disadvantaged, control room operators or visitors are expected to use the control room adequate facilities, and access should be provided
- It is recommended that workstation layouts and work regimes allow operators to change their position at the workstation and to move from their workstation from time to time (whilst ensuring that primary control duties or time-critical tasks are not adversely affected)
- Room dimensions and control workstation layout dimensions should take account of the anthropometric dimensions of the user population
- Control room operators using visual displays should not be facing windows unless these windows are a primary information source
- The layout of the control room should ensure that all off-workstation visual displays (e.g. overview displays) necessary for operators' task are visible and readable from all relevant control workstations
- Circulation of control room staff, maintenance staff and all visitors should be achieved with minimum disruption to the work of the control room operators
- Main entrances and exit should not form part of the working visual fields of operators, unless they have specific responsibilities for checking flow of staff into and out of the control room
- Control rooms should provide for future physical expansion and technological innovation
- Operators should not be located close to ventilation system inlets or air vents

#### *Environmental design principles*

- The needs of the control centre operator should be paramount (over technical considerations)
- Levels of illumination and temperature should be adjustable so that operators can optimise their performance and comfort
- Environmental factors invariably work in combination, therefore, should be considered in a holistic way
- Environmental design should be used to counterbalance the detrimental effects of shift work (e.g. use of temperature increases to compensate for diurnal rhythms)
- The design of environmental systems should take account of future change (e.g. equipment heat emission, workstation layouts and work organisation)

Basic physical constraints are imposed by the location of the interface. If such a location is a control room, then the most basic of these are the architectural constraints of the building in which the control room is located. These are considerations that should be made early and could have a detrimental effect on the effectiveness of the control room. For example, there may be architectural constraints on the positioning of consoles (e.g. blast proof walls or channelling under raised floors) which reduce the options concerning the number and type of interfaces that can be housed in the control room. There may be environmental constraints (e.g. the temperature and humidity of the room) affecting suitability of the equipment. The size and shape of the control room will have an impact on the workspace design and consequently on the number and positioning of the VDUs. This will in turn determine how operators will interact together and work as part of an operating team. Social contact should be catered for by grouping operators so that conversation is possible without compromising efficiency, especially important in larger facilities during quieter periods when staffing levels are lower. The interfaces provided in a centralised location such as a control room may be supplemented by interfaces actually on plant or in other locations, to provide information on system status. These may require ruggedisation to allow them to conform to safety requirements for equipment to be located on the plant or factory floor. Other specialised requirements are for seismic protection in certain geographical regions or for protection against shock or vibration on a ship or aircraft.

Consideration should also be given to the type of communication that needs to take place, and how. Physical disturbances, such as noise, are often cited as distractions in operational environments. However, not all noise is a disturbance, and in certain situations, the overhearing element can be an important tool. For example, in a control room situation when dealing with an incident, it may be vital that certain roles are sitting next to each other so that they can interact and can overhear conversations and interactions with other operators. In other situations, specifically in multidisciplinary teams, this type of communication is unwanted, and incorporating extra room between desks in the initial design stages may help to mitigate this. Interestingly, in situations where everyone is carrying out the same task at a similar volume and fairly consistently (e.g. a call centre), then the disturbance is a lot less, and more operators can be situated closer together.

Since environmental factors can cause major problems in a control environment, the greatest value is in offering operators individual control of their own environment. For example, having access to and being able to set their own light and heating levels. In order to mitigate issues such as this, operator participation is key. By involving the operators and identifying how they want the space to work for them and what they need can lead not only to better designs but also to fewer issues down the line.

## OUTLINE SYSTEM DESIGN

### INFORMATION ANALYSIS

The purpose of any display at work is to deliver information to someone to allow them to perform a task, whether this be passive (e.g. monitoring) or active (e.g. calibrating or fault diagnosis). In doing this, display design must eliminate or reduce errors of sensing, recognition, perception and decision; in other words, people should be able to recognise the relevant information against its background, distinguish its meaning to understand what is required and use the information to make decisions and perform tasks. The right information must be communicated in the right form to the right person at the right time.

One useful technique to employ in early screening of display designs can be borrowed from the world of work study – the systematic questioning process, used in examining a method and process analysis. This can be applied mainly to the information content of displays (but to an extent also to their form) and is a systematic look at what information is to be contained in displays before any consideration of how such information is to be presented. The questioning procedure shown

**TABLE 14.4**  
**Questioning Procedure for Information Analysis**

|                                                      |                      |                                     |                                 |
|------------------------------------------------------|----------------------|-------------------------------------|---------------------------------|
| What information is to be displayed?                 | Why is it necessary? | What else could be displayed?       | What should be displayed?       |
| Where is the information to be displayed?            | Why there?           | Where else could it be displayed?   | Where should it be displayed?   |
| When is the information to be accessed/communicated? | Why then?            | When else could it be communicated? | When should it be communicated? |
| Who is to have/use the information?                  | Why them?            | Who else could have/use it?         | Who should have/use it?         |
| How is the information to be presented?              | Why that way?        | How else could it be presented?     | How should it be presented?     |

*Source:* Author's (JRW) own records.

in Table 14.4 allows a systematic and rigorous look at all parts of the system, whether in analysing existing displays or assessing needs for new ones (see also ISO 11064 part 5 section 5). An alternative method that can be used to understand the information requirements in a control context is the consideration of a distributed cognition approach, and the propagation and flow of information within a control context (Cox et al., 2007). Examining past incidents or interesting situations by using a timeline-based method such as the critical decision method (Klein et al., 2002) can be a thorough way investigating distributed cognition.

## OPERATOR DEMANDS AND WORKLOAD

Increasingly, people at work are interacting with machines or processes almost entirely through an interface rather than by direct sensing and physical actions. As a consequence, the interface will be a major determinant of the load on the operator. Broadly speaking, we can talk of physical workload and of mental workload (MWL). In general, we should seek to minimise static or dynamic physical workload imposed on operators by modes of control (e.g. heavy cranks and multiple valve opening) or by equipment layout (e.g. the position of a display giving a worker neck ache to see it or controls requiring reach to awkward positions).

For MWL of process control operators, the position is less clear, mirrored by much debate over the whole notion of MWL and its measurement (see Chapter 18). Despite the theoretical arguments against it, the notion of the arousal curve, indicating lower performance or higher errors when people are at high or low arousal levels, has great face validity for designers and allows general guidance to be given. Stressors from the work (e.g. time pressures), environment (e.g. intrusive noise) and personal factors (e.g. lack of sleep) should be maintained at intermediate levels, certainly not at extremes. Consequences of loads on operators which are too high or too low may be lapses in attention, cognitive lock-up, less coordination and timeliness in performance, shedding of tasks in random fashion, impatience, irritability and so on.

When designing new control contexts, it is valuable to be able to predict or anticipate the level of workload that will be experienced. The interaction of multiple factors in many tasks can make simple prediction of demand difficult however. One approach that has been developed to evaluate both current workstations and predict demand of workstation configuration changes or new systems is the operational demand evaluation checklist (ODEC) (Pickup et al., 2010). This tool used a repertory grid technique (see Chapter 7) to identify the constructs in the work of railway signallers that influenced the levels of demand that were imposed on them. This process elicited a set of elements related to the operational infrastructure (e.g. controlled signals, stations), indicators (e.g. level crossing CCTV), process (e.g. isolations, incidents) and service pattern (e.g. train splits (where a service

### Data Entry

The ODEC application includes both guidance and tooltips for your data entry. The data entry form is shown below.

Click on the question marks to see examples of the help content.

| Guidance | Entity                                   | Data To Collect  | Observed Data           | Workload | Notes                                          |
|----------|------------------------------------------|------------------|-------------------------|----------|------------------------------------------------|
| ?        | Regulating Locations                     | Number of        | 3                       | Medium   |                                                |
| ?        | Point ends                               | Number of        | 55 <input type="text"/> | High     | <input type="text"/>                           |
| ?        | Points traversed in a normal train path  | Number of points | 15                      | High     |                                                |
| ?        | Points traversed in a complex train path | Number of points | 25                      | High     |                                                |
| ?        | Timetabled trains                        | Maximum per hour | 6                       | Low      | This will not be increased under the new plan. |

**FIGURE 14.3** Example of an electronic version of ODEC.

is divided into two)); these elements were then described as constructs. This process resulted in the development of a quantitative tool that enabled the number of different elements present within a control context to be used to inform an estimate of whether high, medium or low workload would be expected; this tool is now actively used to support the redesign of signal control environments in the United Kingdom.

Figure 14.3 shows a small section of ODEC to be filled in by the human factors practitioner. This shows a small section concentrating on the number of trains and how complex the area is. When the observed data are filled in, the workload column updates automatically with low, medium or high. The notes column can be used for any additional observations that may impact the prediction of workload.

However, this rather static technique does not cope well with adaptive technologies: although in more regimented, predictable work environments, such as call centres, workload can be accurately predicted and modelled. As an example, proposed new signalling systems within rail are to have work environments that can be rapidly reconfigured in response to changes in work activities or personnel availability. This will mean that the work can be spread between different operators at different times depending on demand. So in a normal everyday situation, one operator may control a large area and be assisted by automation. However, during an incident, when the workload increases, they may wish to transfer some of their area to another operator. This raises new challenges for workload prediction and requires a multidimensional tool capable of coping with dynamic and varied physical environments (i.e. varied sizes of geographical areas, numbers of points, numbers of signals), changing mental demands (rapidly going from very low workload to very high workload) allowing for different levels of local knowledge and experience (an individual controlling a larger area may have less local knowledge and be more reliant on the decision support system) and supporting the quick and efficient sharing of information.

### CONTROL DYNAMICS CAPABILITY AND REQUIREMENTS

Traditionally, much work within the field of E/HF focused on *active* control processes, where the operator had either continuous or discrete control over an active and directly controlled context. Increasing levels of automation and trends towards combining multiple, local control rooms into larger, remote management centres not only have influenced the types of control interactions but have fundamentally changed the nature of the *control* task as well. For example, an operator who is controlling the movement of a bomb disposal robot might be described as having a traditional continuous control task. Many such technologies now include automatic sensing systems, to avoid collisions, and aid navigation, as well as sets of remote cameras which can be monitored.

Therefore, what was once simply a continuous control task is now a combination of continuous control and knowledge-based decision making that influences strategy and requires assimilation of information from different sources. The operator might be required not only to control robot movement but also to select, for example, display configurations from cameras.

Many models that describe *levels of automation* have been presented (Balfe, 2010; Parasuraman et al., 2000). These models typically consider the extent to which the technology or operator has responsibility for decision making, analysis and execution of actions. Whilst the value of these different levels remains, the complexity and combination of different technologies in many control contexts mean that there is perhaps a need for higher levels of granularity in the descriptions of some of the levels.

Examples of discrete control are switch and valve operation to start up a plant; the latter is when an operator has to keep a plant working with several parameters (e.g. temperatures, flow rates, pressures) held at specified values and when fluctuations or transients mean that continual adjustment of the variables is needed. Continuous control is also exhibited in most kinds of transport, whether a car, boat, helicopter or submarine. Many modern work tasks require a mixture of continuous and discrete control, and interface design must allow for this. PC use will involve both discrete control – for example, the sequence of keyboard initiated steps required to edit text, and also continuous control – for example, use of mouse or space-ball or graphics tablet to input graphics or to *walk through* or around the display. Handwriting is a fundamental form of continuous control.

Amongst the factors to be accounted for in a continuous control interface are the following:

- What order of control is required? Is the operator making simple step (or zero order) inputs to control distance and position, or is the control of rate (first order), acceleration (second order) or of the much higher orders found in complex systems?
- If the operator is tracking a course, is this pursuit tracking (where target and cursor both move) or compensatory tracking (where change is shown in the discrepancy between target and cursor)? The former mode has control advantages, but the latter can make for more effective panel layouts by saving space. Whatever the choice, it will have considerable implications for display design.
- Would any form of assistance for the operator be of value? We may apply control aiding (e.g. control can be of cursor position or rate of movement according to whether large distances must be tracked), or display quickening, preview or prediction (all techniques to help when operators might want advance warning of what is to come or of what effects their actions might have).

For a full discussion of human factors issues to do with continuous control, the reader is referred to Kantowitz and Sorkin (1983), Sanders and McCormick (1992) and Wickens and Hollands (2000).

## MODALITY AND INTERACTION PREFERENCES AND NEEDS

Although controls are most often hand operated and displays are most often visual ones, other modalities are possible. Control may be affected by feet or legs, especially when power must be transmitted, when the hands may be fully occupied and when the control task is one of discrete actuation. Examples are an on–off button for a power press, pedals in most forms of land transport, a foot mouse for a computer or a pedal/shuttle mechanism as on a sewing machine. Novel forms of control may be found with computers, including use of eye or head movements, and increasingly the use of gestures, either tracked (e.g. using systems such as Microsoft Kinect) or through capacitive touch screens, are used to interact with systems. Finally, one of the most common forms of input, albeit more for human–human systems, is speech. Although technical and human factors difficulties remain, speech is increasingly used for system input.



**TABLE 14.5****Examples of Influential Factors Affecting Displays in the Workplace and Impact of Factors on Display/Human Factors Considerations**

| <b>Influential Factor</b> |                                                  | <b>Human Factors Consideration</b>                                                                                                                               |
|---------------------------|--------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| User                      | User experience                                  | Extent to which display has been customised; ability of user to select appropriate display settings (e.g. scale, brightness), support for user search strategies |
|                           | User visual characteristics                      | Eyesight (e.g. scale, resolution), colour blindness                                                                                                              |
|                           | User linguistic experience/ability               | Type of text displayed (e.g. character vs. letter based), use of icons vs. text, function available to translate interface                                       |
|                           | User preference                                  | Colour, content (e.g. text vs. icons) selected                                                                                                                   |
| Task                      | Office-based tasks (e.g. word processing, email) | Design of stand-alone peripheral display                                                                                                                         |
|                           | Short tasks (e.g. email, SMS)                    | Use of mobile/tablet devices to complete typing and reading task                                                                                                 |
|                           | Reading/individual viewing                       | Desire to sit/stand in the range of locations (e.g. whilst travelling, on sofa)                                                                                  |
|                           | Shared viewing                                   | Size and resolution of large screen or projected display                                                                                                         |
|                           | Viewing objects or images                        | Resolution and colour display requirements for CAD, video, film or text displays                                                                                 |
| Technology                | Conventional PC monitor                          | Continue to consider standard display screen equipment design considerations, for example, height of screen, angle of screen                                     |
|                           | Laptop                                           | Screen angle, battery use, resolution, contrast                                                                                                                  |
|                           | Mobile device                                    | Environmental conditions (rain, brightness of ambient environment), use of display as interaction device, personal protective equipment (gloves, goggles)        |
|                           | E-paper                                          | Requirements for lighting in ambient environment, resolution, contrast                                                                                           |
| Context                   | Conventional office setting                      | Standard display screen equipment design considerations                                                                                                          |
|                           | Mobile                                           | Varying requirements depending on environmental conditions, postural impact of use of mobile display for extended period of time (e.g. whilst travelling)        |
|                           | Home setting                                     | Use of displays in comfort, prolonged use of <i>occasional</i> work settings                                                                                     |

*Source:* Adapted from Sharples, S., Displays in the workplace, in: *Handbook of Visual Display Technology*, Bristol: Canopus, Springer, 2012, pp. 191–202.

When considering the types of displays that are used in the workplace, Sharples (2012) describes the influential factors associated with users, tasks, technology and context, as seen in Table 14.5.

Displays in the workplace are changing, diversifying and developing. Whilst display quality and power are increasingly improving, developers are building new device types and forms, and users are continually finding new work-related tasks and contexts of use in which to employ them. Device designers and evaluators must ensure that we track these developments carefully and acknowledge the impact of device design and type on user use and interaction (Sharples, 2012).

Selection of mode of input, and especially any decision of whether to use anything other than hand controls, will be on the basis of the task analysis and earlier design decisions. This is true also for display modality selection. Displays may be tactile (e.g. Braille or shape coding of control knobs), proprioceptive (e.g. sensing of correct speed and track for cornering

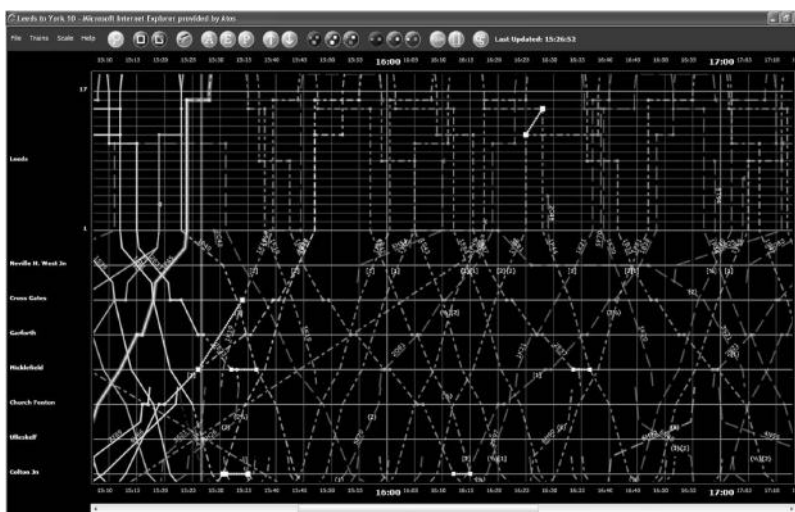
in a car), but most will be visual or auditory with emphasis on the former. In general, an auditory display may be chosen when

- Messages are simple and short (even for speech synthesised displays)
- Immediate action is required
- No later referral to the message is needed
- The information is continually changing
- The message refers to events in time
- Work is in poor viewing conditions – for instance, low illumination or high vibration
- The recipient is moving around
- The recipient is receiving a large amount of other visual information
- The task environment does not have a high degree of background noise

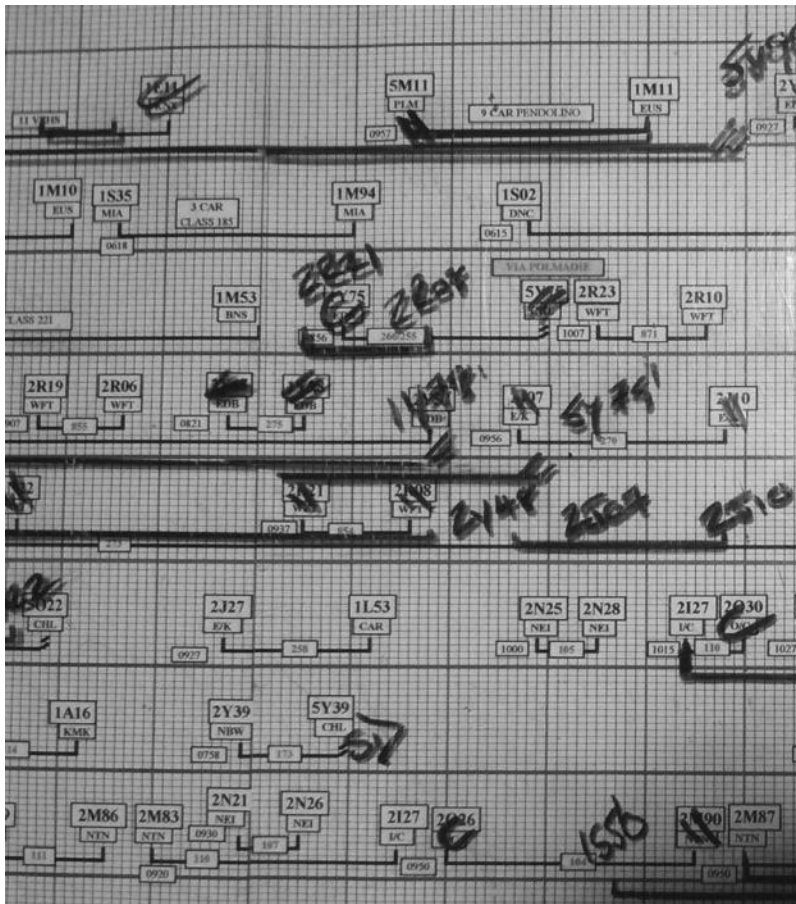
From this, it is clear that a principal use of auditory displays is in warnings or other short signals for action. However, the vast majority of interaction with systems though will be via visual displays, and the remainder of this chapter concentrates on this. The reader should, though, bear in mind the role that auditory signals may be a very important source of information in communication centres. Guidance on the selection and use of input devices in a control room context is again given in ISO 11604 part 5.

The interface design effort must take into account the consequences of events and look at different scenarios and modes of operation and predict the information needs of the operators under these different scenarios. This will include normal operational scenarios and also abnormal conditions, emergencies and maintenance. In some emergency situations, the information for maintaining safe and effective system status may be provided through existing display formats; in other situations, special formats may be necessary. In these situations, interfaces may change in appearance and become cluttered or complicated.

Figure 14.4 shows a train graph. These displays are used throughout the railway industry to help predict future situations and plan for them rather than managing situations reactively. When trains are disrupted or late, this display will start to change. Lines will change colour, and there will be indications in the form of large red dots to alert the operator to a potential timetable clash.



**FIGURE 14.4** An example of a graphical display. A real-time train graph used in rail. (Image courtesy of Atos, Bezons, France.)



**FIGURE 14.5** An example of a station management tool with annotations.

Figure 14.5 shows a tool used to support staff who are allocating trains to platforms at railway stations. This is a large physical tool that operators amend by hand using a china graph pencil (on a Perspex sheet placed over a paper time-based representation of train movements) to reflect any changes to the planned allocations of trains. This physical, shared display has the benefit of always showing an up-to-date version of the plan. It is also extremely easy to make changes and visualise the impact any knock-on effects that those changes will have. These sorts of displays also have the benefit that in the case of an abnormal event, screen-based displays are vulnerable to factors such as loss of power, whereas a physical paper-based representation does not have this vulnerability. For reasons of reliability and safety therefore, physical panel-based displays may be provided as back-up to, or even in place of, computer screens. For example, many control centres have large whiteboards with the entire system pre-drawn on to be used in such emergencies. This is extremely prevalent in rail where VDU-based systems are used.

It is important to ensure that information concerning critical system states is available during emergency situations, to allow the operator to monitor the system and to bring it to safe status, perhaps using an emergency shutdown panel remote from the control room. In an emergency, automatic safety systems may operate and intervention by personnel prevented or restricted. Information provided to the operators is critical to ensure that they know what is happening, that appropriate action is taken and that personnel are able to anticipate future system states. In emergencies, it is imperative that the interface does not add to the workload of staff. Designers of some systems have adopted a strategy to reduce the tendency of people to react in certain inappropriate ways under stress.

by ensuring that user intervention in the system in the event of an emergency is prevented for a set period of time to allow the operators to familiarise themselves with the conditions of the system and with information concerning the failures. This avoids the tendency of operators to try and fit the information they have into familiar diagnoses and solutions, rather than to explore a range of solutions to the problem in hand (this is shown in availability bias and cognitive lock-up – see Hogarth, 1980, pp. 204–234). The famous example is the Three Mile Island incident (Reason, 1990, pp. 189–191, 251), where operators considered that they had correctly diagnosed the origin of failure and took corrective actions based on this. Additional information, which would have indicated that their diagnosis was incorrect, was fitted into the original diagnosis or largely ignored, leading to a delay before the correct cause was identified and mitigating action could be taken.

There are a number of factors that impact on interface design under abnormal operating conditions. These include differences in the tasks to be performed, differences in personnel (including engineers or managers who may be unfamiliar with the displays), differences in workload and thus increased stress from unfamiliar operational conditions. Also, whilst certain industries have incident rooms to handle larger incidents (such as the emergency services) during the early stage of an abnormal scenario, the control room is often used as an emergency control centre, which may change needs for access to consoles or for numbers of people in the room and will certainly increase background noise and distractions.

In the design of screen-based interfaces for complex systems, maintenance is critical and often overlooked. In many cases, errors which bring about system failures actually occur during the maintenance phase of the system life cycle. Maintenance personnel frequently use the same screen display formats (or at least the same system) as operating personnel. Even if the information presentation is not common between these two groups of users, information concerning which equipment is undergoing maintenance (both emergency and planned maintenance) is essential to operating personnel, to assure the safety of personnel and equipment during the operational phases of the system. Although it is usual good practice physically to lock off controls at the panels for equipment undergoing maintenance or repair, it is more difficult to do this reliably working from a screen. Also, in many situations a shutdown may not be possible, for instance, in air traffic control or the emergency services. A further potential problem in safety critical systems, where instrumentation must be checked at intervals as short as 24 h, is that mistakes in checking can bring about an unforced shutdown of the plant.

We cannot cover all relevant aspects of interface specification in this chapter. The available research and guidance on, for example, computer input devices or screen coding is legion. We will have to be selective. There is good coverage in many of the general texts on control rooms and process control identified at the start of the chapter. Future trends in interfaces include the increasing use of augmented reality, and the integration of handheld smart phone or tablet devices for information capture and presentation.

The interface provides the basis for communication between the system and the user. It provides information to the user on current status and performance of the system. In turn, the user can manipulate and monitor the system via the interface. In the design of an interface, the consideration of the information needs of the user(s) is probably the single most important element. The first type of display that will be considered is the *primary* control display, the interface where the operator executes control commands. Within this, we will first of all compare the two display formats of most applicability to control room design – panel displays and VDUs – before looking at some methods of information coding on displays.

One of the principal choices facing the control room interface designer has historically been that between panel and/or VDU displays (or increasingly flat panel LCD displays). To an extent, panels have been replaced by VDU interfaces and flatscreens – so much so that the authors debated whether or not to include this section in the chapter. However, there are good operational (and operator support reasons behind having panels), and many control facility designs still have a mixture of displays from legacy technology, so an idea of the advantages of each and preferred use is still useful.

Panel-based displays are usually wall-mounted panels in modern control rooms, although they can also be found as floor-based consoles. These may be mainly static mimic-type displays, which include dynamic information through LCDs or indicator lights, or they may be projected displays or large interactive screens. They may also be large banks of screens for CCTV monitoring use, such as road traffic control, or monitoring metro systems. For such panel interfaces, the critical decisions are how the information will be displayed (modality) and where it will be located on the control panel. Controls have also traditionally been located on the panel amongst the displays and will usually be allocated to individual displays, but increasingly touch screen technologies allow direct manipulation of displays, either through physical *touching* of screen elements or gesture recognition. One of the most critical factors in the design of panel-based displays is the layout and grouping of information. This is particularly important in CCTV-based applications where lots of images are similar, such as a prison. This must accommodate the full range of tasks that the operator is required to perform. Guidance can be found in ISO 11064-5. It must also accommodate tasks where two or more operators are required to work side by side. For example, information that is required by more than one operator to carry out different functions concurrently will necessitate special consideration for display placement, or information redundancy through display duplication elsewhere on the panel. Alternatively, overview displays can be projected onto a screen on the wall of the control room, to give a large dynamic overview of key parameters, accessible to all personnel. Such shared overview displays can provide support for teams working in a control room environment. However, they raise additional issues for consideration in the design of information, for instance, as follows:

- Is an overview display needed?
- What information should be available on the display?
- How should the information relate to desk-based or individual operator displays?
- Who will operate the configuration of the display?
- What back-up capabilities are available?

The configuration of shared displays has an impact on the design of control room and individual workspace layout.

The other important factor in consideration of the design of panel displays is the relationship between controls and displays located on the interface. The location of the control associated with a particular display or range of displays must clearly indicate which display will provide feedback of the effects of that control action. The movement of the control must also conform to the expectations and stereotypes of the user in terms of the effect it will have on the display (see Wickens, 1992, pp. 324–334).

Control–display compatibility, in terms of the expectations of users and design which fits with common population stereotypes, will depend on displays indicating a response that is expected as a result of the control action taken. Stereotypes may be related to the expectations of a significant proportion of the population – for instance, that a movement of a control in one direction would be expected to change the associated display in a particular way. They may be culturally based – national or professional – for instance, in terms of the meaning ascribed to certain colours. Placement or activation of controls and displays that do not conform to majority expectations or which do not meet principles of compatibility can result in greater training requirements, increased errors or slower reaction or operation times. See Sanders and McCormick (1992) for more information.

Increasing technological complexity has meant that far more demands are placed on the design of the interface to convey relevant and timely information to the operator. The role of the operator is not only one of operation, but through that operation, he or she has a major role to assure the safety of personnel and maintain the integrity of the equipment and the environment. A vast number of data points could potentially be displayed to the operator, and the importance of displaying the right information at the right time is critical. In general, most systems require a dedicated screen to display alarm or emergency information, allowing this critical information always to be available



**TABLE 14.6**  
**Set of Useful Principles for the HCI Design in Control Rooms**

*Principles for control room HCI*

- The human operator must at all times be the highest authority in the human–machine system.
- The operator has to be provided with all the information he needs to accomplish the task.
- The user must be able to maintain a comprehensive and robust mental model of the system and its associated sub-systems (related to situation awareness).
- The system should not ask the operator to input data that are already available within it.
- Information should be presented according to known ergonomic principles so as to ensure that it is conveyed quickly and accurately to users.
- Objects that are to be recognised at various levels of display hierarchy should be designed so they are clearly legible and recognisable at all levels of magnification or zoom.
- Only valid information in terms of time, origin and appropriate resolution should be displayed, and where this is not practical, it should be indicated.
- The information presented by the system should be offered in such a way that it can be easily and unambiguously understood without any additional help.
- The level of attention getting – applied to a particular item of information – should be matched to the importance of that information to the operator and the safety of the system.
- Operators should not be expected to *fly blind*. Controlled objects should be displayed whenever possible (not that this may not be possible with certain systems such as dispatch of resources by radio).
- Interactions with the system should be kept simple, and complex exchanges kept to a minimum.
- They system should aid the operator in inputting information efficiently, correctly and with a minimum risk of errors. For frequent control actions, short cuts should be made available to the operator.
- Appropriate feedback must be provided to the operators at all times.

*Source:* Adapted from Wood (2004).

to the operator regardless of the current system status information that is being displayed at the interface (Bainbridge, 1991). The size of a display and the number of screens actually installed should be considered in the light of anticipated normal and worst-case scenarios, number and types of users, number of data points for display and the control room configuration (see also ISO 11604 parts 4 and 5).

In control rooms, primary operation is often carried out from a VDU console arrangement, with panel displays providing an emergency back-up, this is particularly useful for those with supervisory roles to maintain an understanding of system state, and identify any operators who are experiencing a high level of demand (Table 14.6).

## DETAILED ELEMENT DESIGN

### CONTROL AND DISPLAY SPECIFICATION AND INTEGRATION

We cover a range of display types here, but readers are pointed to ISO 11064-5 for additional information on display specification.

#### Display Formats

Whatever the display type, there are a number of possible formats for the display of process information: including mimic, sequence, alphanumeric, graphical/trend, deviation, object oriented, pictorial, fault and alarm. VDU technologies mean that the information can be displayed in almost any format limited only by the screen size and resolution. Increasingly, displays are developed specific to the system under control, and operators may be allowed their own individual configurations of the workstation. However, whilst some displays still replicate the format



of the local hard-wired equipment display (e.g. a three-point controller), formats such as mimics and graphs are more commonly used.

### **Mimic Displays**

Mimic displays offer a graphical representation of the system. This representation can either reflect the functional relationships between elements of the system (e.g. a schematic representation of process flows) or the geographical/topographical layout of the elements of the system (rail signalling). Whilst research has indicated that neither is more effective when evaluated in terms of operational performance (Vermeulen, 1987), many displays that are to be used for more than one task are based on a functional presentation. Such a presentation allows the sequence of system processes to be followed and can help the operator to visualise functional relationships between system elements, thus assisting with problem-solving tasks. These display types may have a significant impact on the development of the operators' mental models of system function and so need to be designed carefully to ensure that any inferences an operator may make concerning system function are accurate and do not compromise system safety. For example, the geographical nature of signalling systems in rail can assist the operator in establishing a good-quality mental model of how the railway operates on a day-to-day basis. However, the time-based display of the train graph shown in Figure 14.4 enables the operator to visualise how the trains interact with one another in closer proximity and demonstrates how changes can affect other trains.

### **Sequence Displays**

The use of operating procedures to support a wide range of tasks is common practice in many complex systems (but see, for instance, Dekker, 2003). Such procedures may be in the form of a written manual or may be computer based, although procedural tasks which follow a predetermined sequence are increasingly automated, leaving the operator free to perform other operational tasks. (It is well recognised, though, that the trend to automate task elements wherever possible does not always enhance the safety or reliability of the system.) For an automated process sequence, the operator will need to know the stage in the process that has been reached, to allow monitoring and early detection of any abnormal events, anticipation of requirements for manual inputs or operator intervention, and, in the event of a fault, whereabouts in the automated sequence it occurred. Sequence displays can make this information explicit. Where procedural tasks are still the operators' responsibility, then display requirements are to act as a reminder of, and guide through, the sequence of tasks. Sequence displays may take the form of textual lists (see also alphanumeric formats); however, a graphical form such as a flow chart or network diagram may be preferable particularly where the sequences are complex. The displays should provide an overview *map* of the sequence, indicating operator inputs and giving a positive indication of progress through the sequence.

### **Deviation Displays**

These are used primarily for monitoring and fault detection tasks. The displays are used to indicate when a variable deviates outside given threshold values and to present the degree of deviation. The display usually takes the form of a horizontal bar chart, with each variable being represented by a bar and the height of the bar indicating the magnitude of deviation and threshold limits. If the bars are placed centrally on the display, both positive and negative deviations can be indicated. Deviation displays are good for monitoring and fault detection to show movement towards abnormal conditions and to allow a rapid response if required.

### **Alphanumeric Displays**

At the most simple level, alphanumeric displays can either be static (e.g. printed instructions on a display) or dynamic (i.e. the values or the written information change over time). On panel-based displays, information is usually static with the exception of digital numeric displays (e.g. an LED indicating quantitative information). On VDUs, the formats for alphanumeric displays are varied,

from lists to codes to natural language. The literature offers a range of guidelines on formatting and coding (e.g. Helander et al., 1997; Shneiderman, 1998) covering issues such as layout, density, and use of abbreviations. In control room contexts, alphanumeric displays most commonly take the form of lists, static (e.g. process steps for a batch process) or dynamic (e.g. alarm lists) where the values of the variables or information displayed change according to system state. In addition, there are also dynamic displays in which the data change as a result of operator actions rather than process changes initiated by the system, for example entering numerals to adjust a set point or entering default values prior to plant start-up. In practice, there are many hybrid formats where sections of a display (e.g. overlays or windows) are alphanumeric in style.

### Alarms and Warnings

Well-designed alarm displays are critical to support correct decision making and problem solving in the event of system failure or abnormal conditions. There is a wide body of relevant research and guidance, in terms of information format and content (ISO 11064-5), and interaction (e.g. Stanton, 1994). A range of display formats are found in alarms, often hybrid interfaces with hard-wired LED displays or annunciators complemented by dedicated VDU alert lists or highlighted mimics.

### Graphical/Trend Displays

In a primarily panel-based control room, trend and other graphically based information will usually be presented on chart recorders, because such information usually requires the provision of a historical record of plant operation, and so records are often archived for future reference. VDU-based graphic formats are also used for the display of trend information. They offer the advantage that a large amount of historical data can be stored and readily recalled and that the time axis can be readily manipulated. Such displays can be used as a diagnostic aid and as a prediction tool to anticipate future plant states. The train graph used in rail is an example of such a display type (see Figure 14.4) and can assist in planning (Charles et al., 2012). The display of graphs on a VDU is limited by the size and resolution of the display although several graphs can be overlaid for comparison (a maximum of four to six is recommended). Line graphs are recommended for ease of use for most tasks (Figure 14.6).

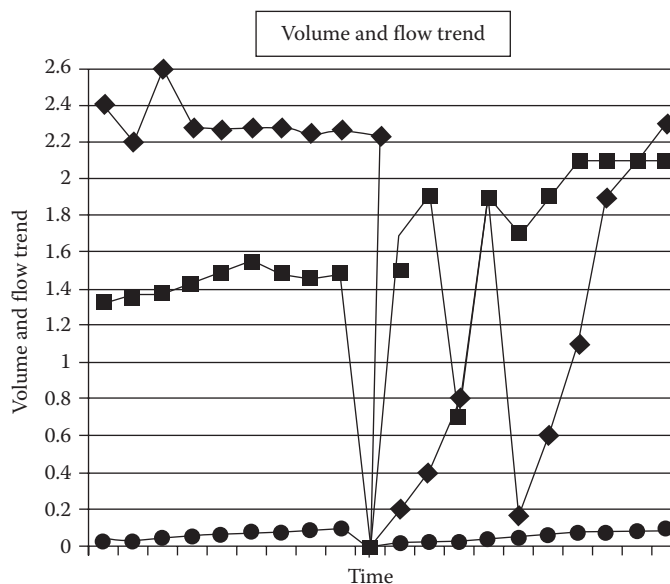


FIGURE 14.6 Example illustration of a trend display.

### Polar Coordinate/Object-Oriented Displays

Polar coordinate or object-oriented displays (also called integrated or shape displays) aim to provide rapid recognition of overall process or system status and can be used as a VDU-based overview display. They commonly take the form of a geometric shape, with the variables to be displayed indicated as points on the geometric object (e.g. at each angle of a hexagon or on the circumference of a circle). The variables are normalised so that when the system is functioning, normally the figure will appear in the expected format, and deviations in the system lead to deviations in the shape. Operators are able to recognise the shapes formed and so will be able to identify quickly any unusual or unexpected change in system status (Figure 14.7).

The displays are based on the principles of Gestalt psychology that people will look for form or shape in information and that processing such information takes up less cognitive capacity than processing the individual data elements. The idea of object-oriented displays was originally proposed by Coekin (1969) and developed by Goodstein (1981) and Wickens (1986). Whilst any geometric figure could be used, studies have proposed polygons, triangles and rectangles.

### Pictorial Display

VDUs can offer options for graphics displays that are not possible with panel-based formats, for example high-fidelity simulations. This flexibility means that pictorial displays and other high-resolution graphics can be used to present information in a way that is easily assimilated and used by the operator. The main advantage of such displays is that they are not constrained by convention and can be tailored to suit the task. As an example, in a paper production process plant, a three-dimensional graphic was introduced to show process flows; operators found it easier to see the impact of problems on the overall continuous process than by using the more conventional two-dimensional mimic displays. Other applications where such displays might be useful include provision of a general qualitative overview of a process which can be backed up by detailed quantitative information, a pictorial representation of a particular item of process equipment in a complex plant, which assists the operator to relate operational problems and maintenance issues to the equipment on plant, and generally any information which cannot be easily represented using pre-defined display formats.

We are increasingly seeing the use of close-circuit television (CCTV) for many monitoring applications, monitoring valves or vessels for steam or gas leaks, observation of personnel working in potentially hazardous areas or in an emergency when access is risky, identifying the location of system problems, or aiding communication in noisy environments or any other situation where there are advantages in giving body signals. Increased use of CCTV will present a new set of ergonomics questions to do with detection, diagnosis, display system navigation, optimum positioning and panning, numbers which can be monitored, operator workload and control room environments.

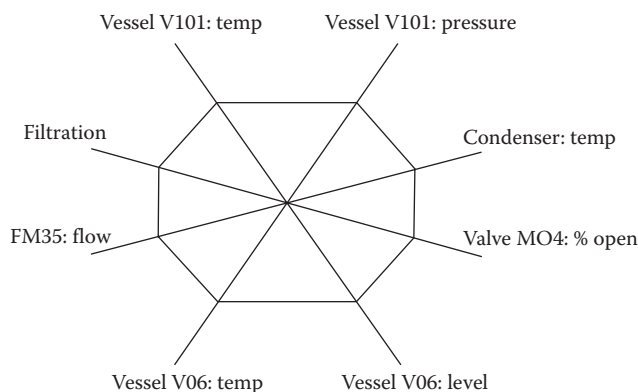


FIGURE 14.7 Detail from object-orientated display format.

## PROTOTYPING AND EVALUATION

Finally, having carried out our initial analysis and conceptual and detailed design, we must evaluate the resulting interface or maybe even the whole control room. However, different approaches will be required depending on the stage of the process: formative assessment will require a different approach to summative assessment. Evaluations will also be carried out independent of design, for instance, as an input to safety audits. Many of the techniques discussed in other chapters of this book must be applied here. We present a set of guidelines written as a checklist, to guide any first-level expert evaluation (Table 14.7). These guidelines are particular to displays in control rooms, and largely VDUs, but many parts would be relevant in other contexts also. The evaluation checklist is sectioned by topic and is merely a simple guide to the issues involved. There is inevitably some overlap between sections since the whole checklist may not always need to be used.

## INTEGRATION AND EVALUATION

### INTEGRATION INTO CONTROL CONTEXT

With the design stage completed, the final stage is to integrate the new or updated technology into the control environment. With the exception of designing a brand-new control facility, new technology will often have to be integrated into existing environments. Good human factors integration is vital throughout the whole design process, in order to increase the effectiveness of the proposed technology and to minimise added costs going into deployment. The development of a human factors integration plan (HFIP) should form part of the initial stages of any large project involving a multidisciplinary team. The benefits of such an approach have been well documented (Cullen, 2007; Wilson and Norris, 2005) across various high-hazard industries. The HFIP provides a structured approach to ensuring human factors issues are addressed throughout the project life cycle and ensures any gaps are handled as soon as possible.

One important consideration, and one that should be built in to any HFIP, is that any design or addition should take into consideration any possible future alterations. Taking rail as an example, the control centres being designed and used today have been designed to accommodate the technology and workforce for the next 20+ years. Consideration should be given to potential technological advances as well as potential automation. In the case of many control centres, new technology is often introduced alongside existing technology. Problems occur when new technology has not been accounted for and may result in major disruption. For example, if the new screens and instrumentation is required to be placed on a workstation, the upgrade is less disruptive and cheaper if expansion has been allowed for.

The main issue faced when introducing new or replacement technologies into established environments is whether it will be willingly accepted and effectively used by users (Venkatesh et al., 2003), especially when the current systems are well used. There is much literature within the information systems field that aims to measure and predict the acceptance of new technology within the context of a working environment using quantitative methods. One of the most well known, validated (King and He, 2006) and replicated of these is the technology acceptance model (Davis, 1989). This has been adapted and added to in the past 20 years (see Venkatesh et al., 2003 for a review), but the basis remains the same: the uptake and acceptance of any technology can be determined by perceived usefulness and perceived ease of use, with perceived usefulness found to have the greatest influence (Davis, 1989). If users perceive the technology to be useful, they will use it! Another advantage of introducing ergonomics early on in a project and adopting an end-user approach is that small technical issues that can frustrate users and lead to poor uptake and acceptance have a high chance of being resolved if subject matter experts are utilised throughout the design process. Another way of ensuring a smooth integration process is to introduce *champions*. These champions can drive the adoption rate and uptake from the front line. This can be a useful approach and one

**TABLE 14.7**  
**Checklist for Evaluation of Control Room Interfaces**

**Displays Structure**

- Maintain a consistent relationship in the way in which all displays are structured throughout the system.
- Avoid the user needing to maintain anything other than simple items in memory when moving from one display to another.
- Ensure that the display structure is transparent to the user, that is, can the user locate any given item of information within the structure, and is the interrelationship between different display pages evident?
- The operator should be able to apply a consistent set of rules for navigation throughout the system.
- Navigation between pages should be simple, and the user should be able to enter and exit the structure at any point.
- Ensure the structure of any hierarchical system is appropriate to the user, for example broad and shallow for expert and frequent users, deeper for novices.
- Ensure the paging structure corresponds to the operator's mental model of the system (or that a coherent mental model can be established through use).
- Provide a page to give an overview of the paging structure.

**Information Structure**

*Navigation*

- The organisation of the display structure should be transparent.
- Moving between screens frequently used in conjunction with each other should be as simple as possible.
- Frequently used displays should be directly accessible (e.g. by dedicated keys).
- Each display format should be labelled with a unique identifier which also shows its place in the navigational structure.

*Division and partitioning of information pages*

- The operator should be able to perform tasks without having to carry significant information from one display to another in memory.
- Information concerning variables which interact with one another should not be divided across display pages.
- Fields for the presentation of particular types of information (e.g. menus, display titles) should be consistent throughout all displays.
- Simple variable relationships can be divided across pages, but there should be a repeat of information to ensure the relationship is clear.
- Preferred display density is dependent on a range of issues such as the format of information, types of coding, number of dynamic data points, display screen resolution and frequency of use.

*Formatting information on the screen*

- Displayed information should be standardised in location of particular kinds of information.
- Fields or screen positions for certain types of information, for example titles, menus, commands and input fields, should be consistent between displays.
- Where possible, displays should be symmetrically balanced.
- Important information should be positioned in the upper left/central/upper right areas of the screen.
- Position and layout of information should provide the user with extra coding about the nature of the information in different parts of the display.
- Data should be grouped to assist with the tasks the user has to perform. This may mean that groupings are on a task-based or functional level.
- Items may be grouped or arranged in a variety of ways dependent on user requirements. For example, criticality – important items are placed prominently and together; frequency – frequently used information is placed prominently on the screen, data frequently used together are displayed together; function – items can be grouped based on function when sequence and frequency are not important; sequence – items are displayed in the order in which they occur, for example process flow.
- The use of too many windows or partitions should be avoided.

(Continued)

**TABLE 14.7 (Continued)**  
**Checklist for Evaluation of Control Room Interfaces**

**Display Formats**

- The formats used should take into full account the task(s) they are used to perform.
- Formats should be matched to user attributes, for example polar coordinate displays make use of operators' pattern recognition capabilities, for rapid detection of system deviations.
- Information should be presented in a format which is consistent with the task requirements of the user (e.g. if qualitative state information is required, then a digital readout may not assist the user and may actually detract from task performance).

**Coding**

*General coding*

- Codes should be used to make information more easily assimilated and to structure displays. In general codes should be consistent – both within the display system and with other codes used on plant or in the process; unambiguous – items in the code should not be confusable; unique – any coding scheme should be clearly distinguishable from other coding schemes.

*Colour*

- The use of colour is generally subjectively preferred over monochrome, although it does not always give improvements in performance. Colour can be used to enhance the appearance of a display or as a code. Care must be taken in the use of colour as it can increase the potential for human error.
- Colour should be used only as a redundant code, that is, all items should be distinguishable without colour, using colour only to enhance the coding.
- Consistent colour coding should be used over all interfaces, in the control room and on plant.
- For accurate discrimination of colours in a code, a maximum of seven should be used.
- Colour can be used to relate items that are similar but separated spatially.

*Highlighting*

- Highlighting should not be overused but used selectively for emphasis and to give feedback.
- Blinking is good for attention getting.
- High brightness has attention-getting properties, but less urgency.
- Reverse video should be used in moderation.
- Underlining to highlight text should be used only if it will not contribute to display clutter and if spatial layout permits.
- Do not overuse different fonts and upper case.

*Labelling*

- Consistency is essential and should be observed in size, font, use of abbreviations and positioning of labels.
- Information content of displays.
- Ensure that a task analysis and systematic identification of information needs have been applied within system design.
- Check that all potential users of the system have been considered in display design, including operators, management, control system engineers and maintenance personnel.
- Ensure that the information content of displays matches the tasks it will be used for (i.e. information required for a particular task should be accessed easily and structured in a way that facilitates task performance).
- Ensure that consistency in the presentation of information is maintained wherever possible between plant and control room.

**Information Content of Displays**

- Ensure that a task analysis and systematic identification of information needs has been applied within system design.
- Check that all potential users of the system have been considered in display design, including operators, management, control system engineers and maintenance personnel.
- Ensure that the information content of displays matches the tasks it will be used for (i.e. information required for a particular task should be accessed easily and structured in a way that facilitates task performance).
- Ensure that consistency in the presentation of information is maintained wherever possible between plant and control room.



that can assist when considering and devising a programme of continuing improvement. It is important when introducing any new technology to consider how this will be used in future, and how it will be updated and improved. Careful thought should be given to developing this programme and ensuring an end-user approach is maintained.

## CONCLUSION

As technology continues to develop, and technical focus continues to change, there is an increasing need to match this growth through adequate control room provision. This chapter has provided an overview of the key steps involved in ensuring a control room is fit for purpose, and importantly, future proof. By ensuring the needs of the operators are considered early in the design process, expensive changes later on may be avoided. By setting out the four main design stages in the chapter, we have aimed to provide the reader with a basic checklist. However, through the examples used and discussed, it is clear to see that these often overlap or are used independently depending on the situation. Control rooms, their uses and needs are not universal, and therefore neither should the human factors integration: a range of tasks need to be supported by the technology and functions within the control room, and it is vital to understand these needs fully before embarking on any design or integration exercise.

## REFERENCES

- Bainbridge, E.A. (1991). Multiplexed VDT display systems: A framework for good practice. In: *Human Computer Interaction and Complex Systems*, G. Weir and J. Alty (eds.), pp. 189–210. London, U.K.: Academic Press.
- Balfe, N. (2010). Appropriate automation of rail signalling systems: A human factors study. PhD thesis, C. Sandom, and R. Harvey, (eds.). Human Factors for Engineers. University of Nottingham, Nottingham, U.K.: Stevenage: Institute of Engineering and Technology, CRC Press.
- Balfe, N., Lowe, E., RafeefAbboud, Y.D., Bye, R. and Murphy, P. (2012). Signaller forms automation: The ultimate irony of automation. In: *Rail Human Factors around the World: Impacts on and of People for Successful Rail Operations*, p. 383. CRC Press.
- Ball, P.W. (ed.) (1991). The guide to reducing human error in process operations. Report SRDA-R3 of The Human Factors in Reliability Group. Warrington, U.K.: The SRD Association.
- Blandford, A., Wong, B.W., Connell, I. and Green, T. (2002). Multiple viewpoints on computer supported team work: A case study on ambulance dispatch. *People and Computers XVI – Memorable Yet Invisible*, pp. 139–156. London, U.K.: Springer.
- Charles, R., Balfe, N., Sharples, S. and Wilson, J. (2012). User trust and acceptance in real time rail planning tools. Paper presented at *The Human Factors of Systems and Technology*. Leeds, U.K.: HFES Europe Chapter.
- Coekin, J. (1969). A versatile presentation of parameters for rapid recognition of total state. In: *Manned System Design*, J. Moraal and K.-F. Kraiss (eds.), pp. 153–179. New York: Plenum.
- COmAH, C. (2011). Buncefield: Why did it happen. HSE Books. Accessed at <http://www.hse.gov.uk/comah/buncefield/buncefield-report.pdf>. March 2014.
- Cox, G., Sharples, S., Stedmon, A. and Wilson, J. (2007). An observation tool to study air traffic control and flightdeck collaboration. *Applied Ergonomics*, 38(4), 425–435.
- Cullen, L. (2007). Human factors integration – bridging the gap between system designers and end-users: A case study. *Safety Science*, 45(5), 621–629.
- Dadashi, N., Wilson, J.R., Sharples, S., Golightly, D. and Clarke, T. (2011). Human factors issues in railway intelligent infrastructure systems. Paper presented at *The Contemporary Ergonomics and Human Factors 2011: Proceedings of the International Conference on Ergonomics & Human Factors 2011*, Stoke Rochford, Lincolnshire, 12–14 April 2011.
- Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
- Dekker, S. (2003). Failure to adapt or adaptations that fail: contrasting models on procedures and safety. *Applied Ergonomics*, 34, 233–238.

- Edwards, E. and Lees, F.P. (eds.) (1974). *The Human Operator in Process Control*. London, U.K.: Taylor & Francis Group.
- Gilmore, W.E., Gertman, D.I. and Blackman, H.S. (1989). *The User-Computer Interface in Process Control: A Human Factors Engineering Handbook*. London: Academic Press.
- Goodstein, L.P. (1981). Discriminative display support for process operators. In: *Human Detection and Diagnosis of System Failures*, J. Rasmussen and W.B. Rouse (eds.), pp. 433–449. New York: Plenum.
- Goodstein, L.P., Andersen, H.B. and Olsen, S.E. (1988). *Tasks, Errors and Mental Models*. London, U.K.: Taylor & Francis Group.
- Helander, M.G., Landauer, T.K. and Prabhu, P.V. (1997). *Handbook of Human-Computer Interaction*, 2nd edn. Amsterdam, the Netherlands: Elsevier.
- Hogarth, R. (1980). *Judgement and Choice*. Chichester, U.K.: John Wiley & Sons.
- Hollnagel, E. and Woods, D.D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton, CRC Press.
- Hubbard, A. and Embrey, D. (2010). Deepwater horizon – summary of critical events, human factors issues and implications. Human Reliability Associates Ltd. <http://www.humanreliability.com/documents/DeepwaterHorizon-HumanFactorsIssuesOG.pdf>. Accessed March 2014.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Hutchins, E. and Klausen, T. (1996). Distributed cognition in an airline cockpit. In: *Cognition and Communication at Work*, Y. Engeström and D. Middleton (eds.), pp. 15–34. Cambridge, U.K.: Cambridge University Press.
- International Instrument Users Association (1998). SIREP-WIB-EXERA report M2656X98. Ergonomics in process control rooms. Part 2: Design guidelines.
- Ivergard, T. and Hunt, B. (2008). *Handbook of Control Room Design and Ergonomics: A Perspective for the Future*. Boca Raton, FL: CRC Press.
- Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge, U.K.: Cambridge University Press.
- Kahneman, D. and Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341.
- Kantowitz, B.H. and Sorkin, R.D. (1983). *Human Factors: Understanding People-System Relationships*. New York: John Wiley.
- King, W.R. and He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management*, 43(6), 740–755.
- Kinkade, R.G. and Anderson, J. (eds.) (1984). Human factors guide for nuclear power plant control room development, EPRI report NP-3659. Palo Alto, CA: Electric Power Research Institute.
- Kirwan, B. (1994). *A Guide to Practical Human Reliability Assessment*. London, U.K.: Taylor & Francis Group.
- Klein, G., Calderwood, R. and Macgregor, D. (2002). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, 19(3), 462–472.
- Kletz, T. (1991). *An Engineer's View of Error*. Ruby, U.K.: Institution of Chemical Engineers.
- Moray, N. (1997). Human factors in process control. In: *Handbook of Human Factors*, G. Salvendy (ed.), pp. 1944–1969. New York: John Wiley.
- Norman, D.A. (1983). Some observations on mental models. In: *Mental Models*, D. Gentner and A. Stevens (eds.), pp. 7–14. Hillsdale, NJ: Erlbaum.
- Noyes, J. and Bransby, M. (2001). *People in Control: Human Factors in Control Room Design*. Stevenage, U.K.: The Institution of Electrical Engineers.
- NUREG-0700 (1996). Human-system interface design review guidelines. Revision 1. Washington, DC: U.S. Nuclear Regulatory Commission.
- NUREG/CR-5908 BNL-NUREG-52333 (O'Hara, J.M., Brown, W.S., Baker, C.C., Welsh, D.L., Granda, T.M. and Vingelis, P.J.) (1994). Advanced human system interface design review guidance. Washington, DC: U.S. Nuclear Regulatory Commission.
- Parasuraman, R., Sheridan, T.B. and Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 30(3), 286–297.
- Payne, S.J. (1988). Methods and mental models in theories of cognitive skill. In: *Artificial Intelligence and Human Learning*, J. Self (ed.). London, U.K.: Chapman and Hall.
- Pickup, L., Wilson, J. and Lowe, E. (2010). The operational demand evaluation checklist (ODEC) of workload for railway signalling. *Applied Ergonomics*, 41(3), 393–402.
- Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction*. Amsterdam, the Netherlands: North Holland.
- Reason, J. (1990). *Human Error*. Cambridge, U.K.: Cambridge University Press.

- Rutherford, A. and Wilson, J.R. (1991). Searching for the mental model in human-machine systems. In: *Models in the Mind: Perspectives, Theory and Application*, Y. Rogers, A. Rutherford and P. Bibby (eds.), pp. 195–223. London, U.K.: Academic Press.
- Sanders, M.S. and McCormick, E.J. (1992). *Human Factors in Engineering and Design*, 7th edn. New York: McGraw-Hill.
- Sharples, S. (2012). Displays in the workplace. In: *Handbook of Visual Display Technology*, J. Chen, W. Cranton, and M. Fihn, (eds.), pp. 191–202. Bristol: Canopus, Springer.
- Sheridan, T.B. (1987). Supervisory control. In: *Handbook of Human Factors*, G. Salvendy (ed.), pp. 1243–1268. New York: John Wiley & Sons.
- Sheridan, T.B. (2002). *Humans and Automation: System Design and Research Issues*. Santa Monica, CA: John Wiley/HFES.
- Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 3rd edn. New York: Addison-Wesley.
- Stanton, N. (ed.) (1994). *Human Factors in Alarm Design*. London, U.K.: Taylor & Francis Group.
- U.S. Department of Defence (1999). Design criteria standard – human engineering, MIL-STD 1472F.
- Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003). User acceptance of information technology: Toward a unified view. *Management Information Systems Quarterly*, 27(3), 5.
- Vermeulen, J. (1987). Effects of functional or topographically presented process schemes on operator performance. *Human Factors*, 29, 383–395.
- Vicente, K.J. (1999). *Cognitive Work Analysis*. London, U.K.: Lawrence Erlbaum.
- Wickens, C.D. (1986). ‘The object display’: Principles and a review of experimental findings, Technical Report No. CPL 86-6. Champaign-Urbana, IL: University of Illinois.
- Wickens, C.D. (1992). *Engineering Psychology and Human Performance*, 2nd edn. New York: HarperCollins.
- Wickens, C.D. and Hollands, J.G. (2000). *Engineering Psychology and Human Performance*, 3rd edn. New York: HarperCollins.
- Wilson, J.R. and Norris, B.J. (2005). Rail human factors: Past, present and future. *Applied Ergonomics*, 36(6), 649–660.
- Wilson, J.R. and Rutherford, A. (1989). Mental models: Theory and application in human factors. *Human Factors*, 31, 617–634.
- Wirstad, J. (1988). On knowledge structures for process operators. In: *Tasks, Errors and Mental Models*, L. Goodstein, H. Andersen and S. Olsen (eds.), Chap. 3. London, U.K.: Taylor & Francis Group.
- Wood, J. (1994). The developing international standard on control room ergonomics, ISO 11064. In: *Proceedings of the 12th Congress of the International Ergonomics Association*, Toronto, Ontario, Canada, 15–19 August 1994.
- Wood, J. (2004). Control room design. *Human Factors for Engineers*, C. Sandom, and R. Harvey, (eds.), pp. 203–233. Stevenage: Institute of Engineering and Technology, CRC Press.



# Section III

---

## *Assessment of Impact of Work on Individuals*

The impact of work on individuals has consequences not only for the design of work but also on their attitudes towards work and tasks, and thus can affect work satisfaction and culture. This section contains eight chapters which cover the ways in which we might measure the physical and cognitive effects of work and tasks on individuals. Although this section is described as being about *work*, the measures extend to interactions in the home or for leisure activities (and, as noted elsewhere in this book, the homework boundary is increasingly blurred anyway).

In most cases, the chapters within this book present a combination of theory and method. If we are understanding the relationship between a task and its impact, there are normally theoretical models that build on physiological and psychological knowledge to explain the generalisable relationship between an activity or interaction of an individual, and the consequences on that individual. Therefore, an understanding of those underpinning theories is essential if data from methods are to be appropriately interpreted.

Most chapters also present a combination of methods that enable some form of direct measurement of an effect on an individual (such as a physiological measurement), some capture information about the task being undertaken (e.g. the weight being lifted or the complexity of the task being completed) and methods of indirect measurement that rely on participant reports of their experience of the task. In most cases, it is important that a combination of these types of measures is applied, as all are important in helping to make decisions about changes to workplace, product or system design. In some cases, the information from these different types might differ, and it is important to understand why this difference has occurred. For example, if someone does not report that they are finding a task more difficult to do but the analysis of heart rate (HR) variability implies that they are exerting generally more effort, does this mean that the methods vary in their sensitivity (in crude terms, one method's data are *wrong* and one method's data are *right*), or does it mean that the participant is indeed exerting more effort but either does not wish to report this (as noted in the *macho effect* described in Chapter 4) or does not perceive this increased effort? The chapters within this section aim to provide the reader with a deep enough understanding of the methods, their limitations and capabilities, to help the ergonomics/human factors (E/HF) researcher to interpret and use data obtained from these methods in an appropriate manner.

Chapter 15 describes the underlying theoretical knowledge related to physiological response to work. They describe response to static and dynamic work and consider the extent to which physiological and subjective measures are able to capture responses to physical work. They note the difference between responses of different people within the population and consider how, in the future, we should assess and design physical and sedentary work for the entire working population.

Compare tools to capture work tasks in holistic way, comparing a range of analytical tools that aim to capture physical work in a holistic manner. MAC, RULA and ART plus whole-body assessments provide detailed descriptions of their value in different types of assessments. This chapter is complemented by Chapter 17, which outlines the underlying biomechanical assumptions, which, along with the physiological bases described in Chapter 16, informs the tools applied to assess physical job demands.

Chapters 18 to 20 each considers different psychological aspects of work. Chapter 18 tackles the much-discussed concept of mental workload. Sharples and Megaw discuss the interaction between physical and mental aspects of work and often use the more general term *workload* within the chapter to note the interactive nature of the different physical and mental demands that people are often exposed to in a work context. This chapter retains the structure used in previous volumes of presenting the theoretical concepts associated with workload, followed by the structured description of different methods used in its measurement and assessment. The authors particularly note some of the theoretical challenges and debates around the different concepts of workload and advocate a multi-method approach to ensure that all aspects of experienced load are captured, whether they are subjective perceptions of that load or measures of physiological response.

Similarly, Chapter 19 presents the theoretical models proposed to support the notion of situation awareness (SA) and describes some of the different measures that have been developed to capture SA. To an even greater extent than mental workload, SA is a topic of much controversy, with some disputing its existence as an independent concept, arguing that it is simply a label for a collection of already established factors, such as workload, attention and performance. Whether this is or is not the case, it is foolish for us to ignore the fact that a concept that attracts the attention of those who are designing and implementing work systems, such as SA, workload or stress (as considered in Chapter 20), is one that enables us to ensure that more account is taken of E/HF requirements in work or system design. The focus of Chapter 20 is on describing the notion of stress in a workplace context and providing a methodology to allow the experience of stress in a workplace to be effectively captured.

The final two chapters in this section focus on physiological measurement. Both Chapters 21 and 22 provide detailed technical, practical and theoretical bases to direct physiological measurement of work interactions and performance. These chapters will provide a useful reference to any students or practitioners considering using physiological measurement in a research or practical context. The increasing affordability and accessibility of physiological measurement can make such tools very tempting, especially when we consider the high face validity afforded by some of the graphical representations, it is possible to create from the collected data. Therefore, it is important that users of these tools not only appreciate the underpinning theory that explains the biological elements that are being captured but also understand the granularity of data collected and reliability of the instrumentation they are using.

In conclusion, if there were one theme which connects all the analytical tools and approaches presented in this section, it is that the measurement and the assessment of consequences of work activity – in terms of the effects on individuals – are made doubly difficult because of problems in defining, in unambiguous terms, the phenomena concerned. We cannot promise to have resolved all of the debates surrounding these issues (and the world of E/HF would be a much poorer place without such debates anyway!) but hope that the reader feels more equipped to make an informed decision about how they should measure the impact of work on individuals, whether using self-report measures, performance measures, analytical tools, physiological measures or, as is most likely, a combination of some or all of these approaches.

**Sarah Sharples**



---

# 15 Physiological Bases of Work Assessment

*Karen Søgaard and Gisela Sjøgaard*

## CONTENTS

|                                                                             |     |
|-----------------------------------------------------------------------------|-----|
| Introduction: Domains of Physical Activity and Muscular Work .....          | 419 |
| Occupational Work Requirements .....                                        | 421 |
| Physiology of Muscle Contraction and Mechanical Energy Quantification ..... | 422 |
| Muscle Contraction .....                                                    | 422 |
| Static Muscle Work .....                                                    | 424 |
| Dynamic Muscle Work .....                                                   | 425 |
| Size of Active Muscle Mass .....                                            | 425 |
| Energy Supply: Cardiorespiratory Contribution .....                         | 426 |
| Cardiorespiratory System at Work .....                                      | 427 |
| Body Temperature .....                                                      | 428 |
| Assessment of Work Capacity .....                                           | 429 |
| Anthropometry .....                                                         | 430 |
| Relative Muscle Load .....                                                  | 430 |
| Relative Metabolic Load .....                                               | 434 |
| Work Requirement/Capacity Ratio: Health Outcomes .....                      | 435 |
| Perceived Workload .....                                                    | 436 |
| Muscle Work and Perceived Load during Monotonous Work .....                 | 437 |
| Computer Mouse Work Case .....                                              | 437 |
| Perceived Load in Physically Heavy Work .....                               | 438 |
| Contemporary Perspectives in Ergonomics/Human Factors .....                 | 439 |
| Combining Improved Ergonomics and Tailored Physical Exercise .....          | 439 |
| Physiological Work Assessment as a Premise for a Sustained Workforce .....  | 441 |
| References .....                                                            | 444 |

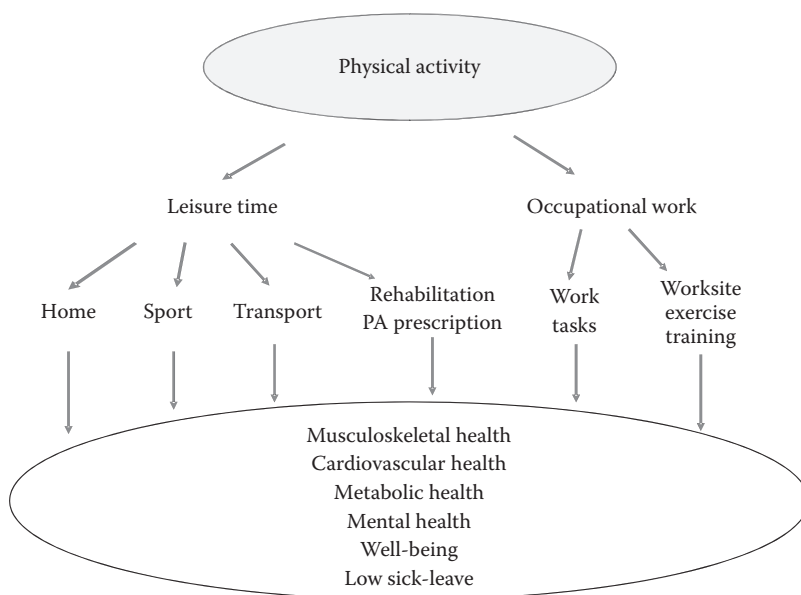
## INTRODUCTION: DOMAINS OF PHYSICAL ACTIVITY AND MUSCULAR WORK

Muscle activity is a cornerstone in physical activity during work and leisure, and is often the greatest stress that the body encounters in the course of daily life. Skeletal muscle accounts for about 30%–50% of total body mass. When skeletal muscle is activated by the nervous system to generate mechanical energy, that is, force or work, such as is experienced in almost every type of task that an ergonomics/human factors (E/HF) specialist might be required to analyse or design, the body must make rapid integrated adjustments from the cell level to the organ systems and coordinate responses within the body. These responses are vital for the delivery of chemical energy for the conversion to mechanical energy that can exclusively occur in muscle tissue. During the resulting energy turnover (i.e. breakdown of chemical energy), thermal energy – or heat – is produced alongside with the mechanical energy, and therefore at the same time as executing muscle activity, bodily responses must ensure heat dissipation.

The regulatory processes that the human body invokes in response to work include increase in HR and stroke volume to increase cardiac output. This will in combination with vasodilation of capillaries in the contracting muscle ensure adequate blood flow with chemicals – in terms of substrates including oxygen. For oxygenation of the blood, it is essential that lung ventilation is increased to balance the oxygen demand by the muscles. Perfusion of the skin in combination with sweating will prevent the overheating of the body that otherwise could be a result of the increased thermal energy produced with the development of mechanical energy by the muscles.

The delicate regulation of these processes and their interplay is entirely dependent on the type of physical activity performed. Typically, physical activity is divided into the domains of work and leisure (Figure 15.1). Within the leisure domain, physical activity may be subdivided into domains of work at home, sport or training, transportation and prescription for rehabilitation. At the work-site, physical activity may be divided into actual work task–related activities, and more recently also worksite exercise training programmes, that have become increasingly popular in many companies, in particular among those with sedentary work. Interestingly, within these time-wise domains during a week, the pattern of physical activity is quite different in terms of the basic characteristics of muscular activity: contraction intensity and contraction mode in terms of static (continuous or intermittent) and dynamic (concentric or eccentric). Further, the volume of active muscle mass, the duration of activity and its repetitiveness are crucial for the physiological responses that subsequently may have an impact on health.

Nevertheless, for decades, a large body of the literature has considered physical activity to provide health benefits irrespective of the site where the physical activity is performed, that is, it is considered enhancing public health whatever the activity is performed at leisure or during occupational work tasks (Figure 15.1). Actually, one of the first studies presenting evidence for the health-enhancing effect of occupational physical activity was Morris et al. (1953) showing that in the buses of London, the conductors, walking up and down the stairs, suffered less from cardiovascular diseases than the drivers, who were sitting all day. Accordingly, the international recommendations for health-promoting physical activity do not distinguish between occupational and leisure time physical activity (Physical Activity Guidelines Advisory Committee 2008, Garber et al., 2011).



**FIGURE 15.1** Domains of physical activity.

However, in this context, attention has not been paid to the extensive literature documenting occupational physical activity to impair health. Scrutinising the work exposure and the imposed physiological responses to the basic loading patterns of the body is essential for identifying risk indicators for impaired – and improved health (Panel on Musculoskeletal Disorders and Workplace 2001, Sjøgaard, 1997, Sjøgaard and Jensen, 2006).

## OCCUPATIONAL WORK REQUIREMENTS

Occupational physical activity involves work requirements that can be classified in terms of high or low mechanical as well as metabolic load and may constitute a risk for overload of the worker. The simplest classification is based on the knowledge of the typical exposure within a sector, for example office work or heavy industry, or based on the job title, for example computer worker compared to construction worker. Such crude classifications are usually applied as estimation of exposure in large epidemiological studies as a means to estimate risk factors for overload. A more comprehensive risk evaluation should consist of information on the general level of force demands, typical body postures, force demands in specific body regions as well as an evaluation of the metabolic load in terms of either HR or oxygen uptake. In addition to the force level, work requirements are defined by the number of repetitions and the total duration of work. Work is classified as highly repetitive if the work cycle time is less than 30 s or if the same movement is repeated more than 50% of the work cycle time. Such type of work is characterised by a stereotyped movement pattern loading the same muscle group again and again. With a repetition of 1 contraction/s, this adds up to more than 25,000 movements in a workday. Such high frequencies have been registered for wrist movements in parcel workers while finger movements in computer work are even higher with about 20,000 keystrokes/h.

While both high force demands and high repetition alone may increase the risk for musculoskeletal disorders, the combination of high force and repetition exponentially increases such risk. This has been shown for both carpal tunnel syndrome and wrist disorders where the combination increases the risk with factors of 14 and 32, respectively (Silverstein et al., 1986 and Armstrong et al., 1985).

The external load and body posture are decisive for the internal muscle force demand. External force demands may constitute risk factors both as peak force, high repetitive force and sustained force. Particularly during lifting, pushing and pulling peak loads may happen. Typically, peak loads may occur as either extremely heavy single lift or due to unexpected sudden loading causing force demand to exceed tissue tolerance level and consequently leading to mechanical damage. This may constitute risk factors for job sectors such as health care worker during patient transfers or construction workers with heavy objects suddenly slipping or sliding. Physically heavy work involving many lifts during the workday is typically found in industrial work, agriculture, slaughterhouses and fishing industry. However, also loading at a lower force level may pose a risk if sustained or repeated for long durations. This is typically a risk factor within light industry, cleaning or kitchen work.

Physically heavy manual work involves a high metabolic rate, especially if all workday is spent walking, standing or involving stair climbing. The metabolic load can be estimated based on HR and oxygen consumption, and in the evaluation, the pattern over the full workday should be considered. Compared to levels during leisure time physical activity, the level of HR or oxygen consumption at work are only moderate but may still pose a risk of overload if sustained for a full workday.

Considering all the aspects of exposure, the different occupational sectors and job types can be classified and described in accordance with the occurrence of factors known to increase the risk for overload as shown in the matrix in Table 15.1. Even though the profiles are not directly comparable, the sectors do represent a continuum ranging from low to high physical load.

In order to protect workers from overload, guidelines and regulations have been introduced. However, a number of important considerations and obstacles exist regarding the suggestion of safe ergonomic guidelines. Therefore, a thorough quantification of work demands is of major importance. Classical risk factors, where guidelines exist, are heavy lifting, high metabolic demands and highly repetitive work. Lifting guidelines usually regulate only the external load, while the actual

**TABLE 15.1**  
**Job Exposure Profiles**

| Sector                   | Job Examples          | General Levels of Force Demands | Posture           | Specified Force Demands in Body Regions        |               |          | Metabolic Load (HR, VO <sub>2</sub> ) |
|--------------------------|-----------------------|---------------------------------|-------------------|------------------------------------------------|---------------|----------|---------------------------------------|
|                          |                       |                                 |                   | S = Static Load, P = Peak Load, R = Repetitive |               |          |                                       |
|                          |                       |                                 |                   | Arm/Hand                                       | Neck/Shoulder | Low Back |                                       |
| Office                   | Computer work         | Low                             | Sitting           | S + R                                          | S             |          | Low                                   |
| Financial Administration |                       |                                 |                   |                                                |               |          |                                       |
| Light industry           | Manufacturing workers | Moderate                        | Sitting           | S + R                                          | S + R         |          | Low                                   |
| Service                  | Cleaners              | Moderate                        | Walking           | S + R                                          | S + R         | S        | Moderate                              |
| Heavy industries         | Production workers    | High                            | Standing          | R                                              | R + P         | P        | Low                                   |
| Health care              | Health care workers   | High                            | Walking           |                                                | P             | P        | Moderate                              |
| Construction             | Construction workers  | High                            | Walking/ kneeling | P + R                                          | P             | P        | High                                  |
| Military, police         | Fighter pilots        | Ultra high                      | Sitting           | S + P                                          | S + P         |          | Low                                   |

load on joints, tendons and muscles cannot be evaluated without data regarding body posture and the capacity of the person. First of all, regulations often approach the size of the external burden, but this is easily applied only in lifting when the weight of the burden can be estimated. For pushing, pulling or sustained holding, force estimation is much more difficult and often must be estimated based on measurements of muscle activity.

In addition to the external force requirements, also high demands on coordination and precision may add to joint and muscle loading due to large amount of co-contraction in order to stabilise the joint. Finally, the work task must be described in terms of muscle activation, for example if muscle contractions are static or dynamic, sustained or intermittent, number of the activated muscles, if they are small or large and if force level is high, low, fluctuating or sustained for long duration.

## PHYSIOLOGY OF MUSCLE CONTRACTION AND MECHANICAL ENERGY QUANTIFICATION

The peak load, the static load, the repetitive workload and the accumulated load can all be quantified in mechanical energy units according to Newtonian physics (Table 15.2). High peak loads may challenge tissue mechanical strain while repetitive contractions may challenge tissue metabolic load. Muscle force may be performed as static contractions (no movement of body segments and with the muscle at constant length, also termed isometric) or as dynamic contractions. In the latter case, the muscle will change length by either active shortening (concentric contraction) or active lengthening (eccentric contraction). In voluntary dynamic work, active skeletal muscles contract rhythmically (shorten and lengthen). The shortening phase is always active, while that of lengthening may be passive during relaxation of the muscle or active when resisting external forces.

### MUSCLE CONTRACTION

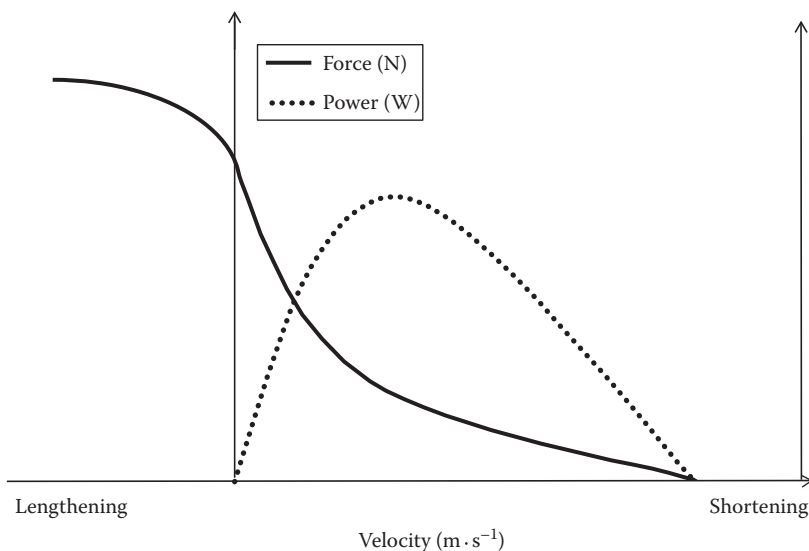
Voluntary muscle activity is initiated in the central nervous system and executed via impulses from the motor cortex to the peripheral motor nerves resulting in electrical signals to propagate along

**TABLE 15.2**  
**Newtonian Exposure**

|                                                                    |
|--------------------------------------------------------------------|
| • Peak                                                             |
| Per contraction                                                    |
| <i>Dynamic work</i> : force * distance (N m = J)                   |
| <i>Static 'work'</i> : peak force (N) or force * time (Ns)         |
| • Mean                                                             |
| Work intensity                                                     |
| <i>Dynamic work</i> : force * distance * frequency (J/s = W)       |
| <i>Static 'work'</i> : force * time * frequency (N, mean force)    |
| • Accumulated                                                      |
| Total amount of work                                               |
| <i>Dynamic work</i> : force * distance * freq * duration (N m = J) |
| <i>Static 'work'</i> : force * time * frequency * duration (N s)   |

the muscle. These can be monitored in terms of electromyogram (EMG), which can give information on contraction intensity from its amplitude. Also the timing of the contribution by various muscle groups in a voluntary movement can be identified and work technique improved. Finally, the EMG signals can reveal possible muscle fatigue when increasing its amplitude and/or decreasing its frequency content in a standardised contraction (see Chapter 21 for further discussion of EMG measurement).

Muscle activation results in the development of muscle force and/or movement, and their inter-relationship during maximal activation are depicted in the force–velocity curve (Figure 15.2). This shows that the highest force can be developed during eccentric contraction, and highest tissue strain with risk of muscle–tendon rupture is seen in such conditions. Force decreases with increasing shortening velocity, which implies that with increased work speed, less force can be produced per contraction. Importantly, however, since power is force \* velocity empirical data have shown that muscles can perform the largest power at around one-third of maximal shortening or one-third of maximal force (Figure 15.2). For peak power, also the maximal rate of force development and shortening must be taken into consideration.



**FIGURE 15.2** Force–velocity–power.

Muscle strength is generally measured during static contractions in standardised body postures where the muscle groups in question contract at optimum length (see section 'Assessment of Work Capacity'). This measure termed maximal voluntary contraction (MVC) is a standard reference value when evaluating workers' strength compared to normal population values or when evaluating the relative muscle loading during work. It is measured in Newton (N) and can be compared to the external loads measured in the same unit (1 kg ~ 10 N), for example, during manual material handling such as lifting, pushing, pulling and carrying. For example, when trying with maximum effort to lift a load by the elbow flexors, which is too heavy to be lifted by these muscles, they will perform an MVC.

## STATIC MUSCLE WORK

In static work, the muscle contraction produces no visible movement, for example, in a limb. In physical terms, therefore no work is performed, but in physiological terms, the muscle does perform work because it must actively develop force under the conversion of chemical to mechanical energy. Therefore, in work physiology, the concept of static muscle work is introduced and quantified as force \* duration (N \* s) (Table 15.2). As mentioned earlier, the muscle length in a static (isometric) contraction is unchanged, whereas the force output may be either unchanged or change over a period of time varying from a few seconds to several hours. When force is maintained over a longer period of time (minutes or hours), this is termed *sustained static work* and results in physiological responses that are quite distinct from those during *intermittent static work*, where each contraction period is few seconds or less and interspaced with relaxation or resting periods.

Static work increases the pressure inside the muscle in proportion to its force development. In sustained static work, this mechanical pressure causes mechanical compression of blood vessels occluding blood circulation partially or totally to the contracting muscle, thereby significantly impacting on the physiological responses. The delivery of nutrients and oxygen to the muscle and the removal of metabolic end-products from the muscle are hampered. Thus, in static sustained work, the muscle fatigues more easily than in intermittent static work where repeated resting periods occur that allow blood to flow through the muscle and ensure homeostasis and dissipate heat from the energy turnover processes (Sjøgaard et al., 2004). In that sense, intermittent static work resembles dynamic work as described later.

The most prominent circulatory feature of static work is a rise in blood pressure – a regulatory mechanism opposing the increased pressure inside the muscle in order to maintain a pressure gradient. This will continue up to ~50% MVC to ensure some blood flow to contracting muscle. Beyond this, usually full occlusion of muscle blood flow will occur, but already from around 10% MVC, a partial occlusion occurs, which means that even at such low levels of sustained static contraction, the muscle is insufficiently perfused. HR and cardiac output do not change much, particularly, if the static contraction is submaximal in nature and performed by a limited muscle mass. Above a certain intensity of effort, blood pressure increases in direct relation to the intensity and duration of the effort. Furthermore, at the same relative intensity of effort, static work with large-muscle groups produces a greater blood pressure response compared to work with smaller muscle groups. However, when working with small muscle groups such as, for example, arm muscles in overhead work, then blood pressure increases even more to overcome also the hydrostatic pressure due to the elevated muscles above heart level.

In occupational work, purely static contractions hardly ever occur, the most common being a low-intensity sustained contraction with very small variations both in muscle length and in the strength for stabilising body postures or objects that must be maintained in position. Examples with minimal variation, that is, almost purely static, are common within the medical or dentistry context under circumstances where pressure needs to be applied in a consistent manner for a long time. Another example, of quite static work, is the operation of technologies such as ultrasound probes, microscopes and pipettes.



## DYNAMIC MUSCLE WORK

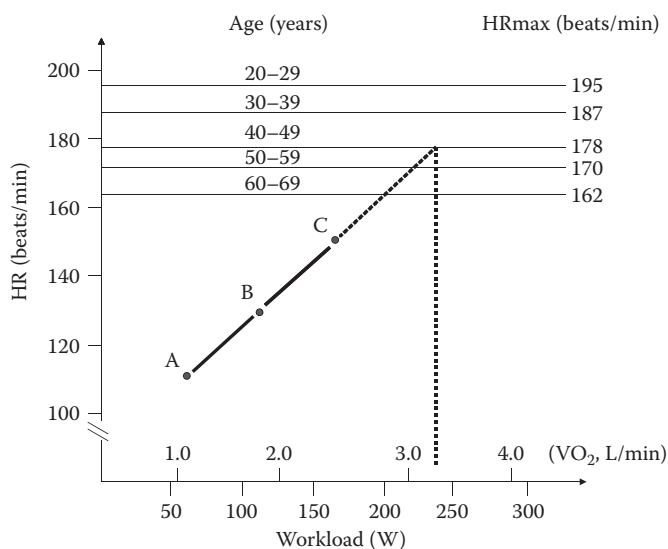
During dynamic work, the muscle shortens during force development and therefore performs work also in physical terms and is measured in  $N \cdot m = J$  (Joule). Likewise, work intensity is measured in Watt ( $W = J/s$ ) (Table 15.2). Dynamic work is characterised by short periods of muscle contraction separated by periods of relaxation or partial relaxation where muscle tissue pressure will be low and allow blood to flow through the tissue. The same is true for intermittent static work, and physiological response to these time-wise dynamic activities is similar although no work in physical terms is performed during intermittent static muscle activity.

The blood flow to the muscles is increased to match metabolic needs as active muscles consume more oxygen and nutrients especially during concentric contraction. The increased blood flow to the active muscles is achieved through increased pumping of the heart (cardiac output), an increased number of open blood vessels in the active muscles and decreased blood flow to inactive areas such as kidneys and liver. HR and oxygen consumption ( $VO_2$ ) in the muscles increase linearly in relation to the amount of active muscle mass and working intensity (Figure 15.3). Also pulmonary ventilation is heightened owing to larger tidal volume in the form of deeper breathing and increased breathing frequency. The purpose of activating the whole cardiorespiratory system is to enhance oxygen delivery to the active muscles. The level of  $VO_2$  measured during dynamic muscle work indicates the intensity of work. The maximal  $VO_2$  indicates a worker's individual cardiorespiratory capacity for dynamic muscle work (Figure 15.3).

## SIZE OF ACTIVE MUSCLE MASS

The physiological responses to muscle work also depend on the size of the active muscle mass. In the case of dynamic work, with smaller active muscle mass (as in the arms and hands), the total  $VO_2$  is smaller than in dynamic work with large muscle masses. At the same external work output, dynamic work with small muscles elicits higher cardiorespiratory responses (e.g. HR, blood pressure) than work with large muscles. The high-speed repetitive type of muscle work with small muscles such as fingers during dynamic handgrip exercise produced almost similar cardiorespiratory responses as static muscle work with the same small muscle groups.

The energy consumption measured as  $VO_2$  can attain higher values in work tasks requiring large muscle groups than in those requiring smaller muscle masses. In young individuals, the highest



**FIGURE 15.3** Heart rate vs. workload and oxygen uptake relationship.

$\text{VO}_2$  attained during a two-arm cranking exercise is about 70% of the  $\text{VO}_2$  during maximal two-leg exercise. The maximal  $\text{VO}_2$  for two-leg exercise declines progressively with age (Figure 15.3). However, during two-arm exercise, and generally during exercises with small muscle masses, the differences between individuals due to age are smaller. At a given submaximal  $\text{VO}_2$ , HR, systolic blood pressure and pulmonary ventilation are higher during arm-work than during legwork or combined arm- and legwork (Aminoff et al., 1998).

The maximal  $\text{VO}_2$  values are usually assessed during dynamic leg exercise with large active muscle mass such as pedalling on a cycle-ergometer, running on a treadmill or stepping up and down on a bench. The relative workloads are often expressed as the percentage of the individual's maximal  $\text{VO}_2$  attained during dynamic exercise with large muscle masses. The maximal  $\text{VO}_2$  for arm-work, however, cannot be estimated from legwork and vice versa. The maximal  $\text{VO}_2$  varies greatly between arm- and legwork due to the large difference in active muscle mass. The external workload that is quite light for legwork may therefore be quite exhausting for arm-work. Hence, to express the relative workload, the assessment of the maximal  $\text{VO}_2$  should be done during muscular exercise that resembles those activities required during actual work. When the workload is related to the maximal  $\text{VO}_2$  attained with the equal amount of active muscle mass, the differences in physiological responses become smaller, but are still higher, for instance, during prolonged arm-work than during prolonged legwork.

In summary, physiological responses are specific in nature and depend on the type of muscle work, the size of active muscle mass and the ratio of dynamic and static work.

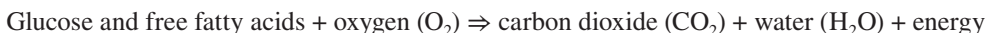
## ENERGY SUPPLY: CARDIORESPIRATORY CONTRIBUTION

In the muscle cell, mechanical energy is produced by the conversion of chemically bound energy. The immediate form of chemical energy requested is adenosine triphosphate (ATP), which will convert to adenosine diphosphate (ADP), heat and mechanical energy in terms of muscle contractions:



Available stores of ATP are very limited and therefore spent within the first few seconds both in dynamic and static heavy muscle contractions. Therefore, ATP continuously has to be built up from energy obtained mainly by the oxidation of glucose and fatty acids. The contribution of energy from protein breakdown can be ignored since the function of protein mainly is to provide material for tissue repair and growth.

The metabolism leading to the release of energy can take place either aerobically (with oxygen) or anaerobically (without oxygen). Oxygen is transported to the muscles by the cardiorespiratory system. If there is enough oxygen available, the more efficient aerobic pathway is followed, because it leads to a more complete metabolism of energy-rich nutrients and allows longer and more efficient work performance due to less fatigue. The form of the aerobic energy production with oxygen is as follows:



In anaerobic muscle work, that is, without oxygen, muscle fatigue occurs rapidly, due to a combination of many metabolic changes in the muscle, ultimately due to insufficient rate of recharging the ATP from ADP. Accumulation of forms of chemical energy such as lactate, hydrogen ions (decreased pH), ADP, potassium and calcium translocations plays a role in the development of fatigue particularly during anaerobic energy turnover. Muscle work always has an anaerobic metabolic component in the following situations:

1. At the onset of muscle activity
2. During heavy dynamic work when the energetic demands exceed 50% of the individual maximal  $\text{VO}_2$
3. During sustained static work exceeding about 10% of the muscle strength

Some of the lactate produced during muscle work can be metabolised in the muscles or be removed by the blood. However, during very heavy dynamic work or during sustained static contraction, blood circulation cannot keep up with the demands for oxygen supply, lactate removal and the general maintenance of homeostasis including tissue temperature. This leads to fatigue and finally exhaustion, that is, the inability of the muscle to maintain the requested force or power. Time to exhaustion is also called endurance time, and this will decrease as the force output of the contraction or the power increases. At the highest force levels or power output, the endurance time is only a few seconds, while it is more than 1 h and with only modest perceived fatigue if the intensity of static contraction is no more than 5% of muscle strength, that is, of a static MVC.

### CARDIORESPIRATORY SYSTEM AT WORK

The cardiorespiratory system, which supplies the voluntary contracting muscles with energy, includes the heart, lungs, blood, blood veins and nutrient storages. During muscle work, the main functions of the cardiorespiratory system are to transport

1. Oxygen ( $O_2$ ) from lungs to muscles with blood flow
2. Glucose and fatty acids from storage in the liver and fatty tissue to active muscles
3. Carbon dioxide ( $CO_2$ ), water ( $H_2O$ ), lactate and other metabolites from muscles to the lungs, liver and kidneys for excretion and metabolism
4. Heat from active muscles to the body surface

In order to meet the demands of an increased muscle activity, the cardiorespiratory system can increase the oxygen-transporting capacity by a factor of 10 or more depending on the physical capacity (see the following equation). The major change is the linear increase in HR with increasing workload and the corresponding oxygen uptake (Figure 15.3). At rest, HR is around 60 beats/min (bpm), and it may increase to maximally 180 bpm for a person around 45 years of age, which is a threefold increase. A younger person has a higher maximal HR and may therefore increase oxygen uptake even more. The stroke volume of the heart will also increase with workload but most likely only by around 20%. Finally, the oxygen extracted in the muscle in terms of the oxygen tension in the arterial blood minus its tension in the venous blood, that is,  $(a - v)$  oxygen extraction, will increase about threefold for a physically active person. Both the increase in stroke volume and muscle oxygen extraction may be further augmented by aerobic training. Thereby, a young well-trained person may overall increase oxygen consumption almost 20-fold. Interestingly, in healthy persons, the capacity of pulmonary ventilation by the lungs seems not to be a limiting factor and has sufficient capacity to fully oxygenate the blood in the lungs even at the highest rate of oxygen uptake.

A 10-fold increased oxygen consumption =  $(3 * HR)(1.2 * \text{stroke volume})(3 * (a - v) \text{ oxygen extraction})$ .

The recovery of muscle function following intense prolonged work may not be complete even though all cardiorespiratory variables have returned to initial levels. After static contractions, HR may normalise, but after repeated contractions, muscle strength and endurance may still be reduced for a prolonged time. After persistent static or dynamic muscle contractions to exhaustion, full recovery may take several hours or even days. The most likely cause of the reduced performance capacity is damage in the structures of the muscles (including ruptures and oedema formation) and maybe also the depletion of energy-rich nutrients in the muscles (Sjøgaard and Søgaard, 1998, Sjøgaard and Søgaard, 2012, Sjøgaard et al., 2000).

Energy turnover during dynamic muscle work is often assessed as  $VO_2$  in litres per minute (L/min) or related to body mass (mL/min/kg) because of the linear relationship between oxygen uptake and energy turnover quantified in Watts, for example, measured in calorie-metre-chambers. On a mixed diet, the caloric value of 1 L oxygen consumed is equivalent to the conversion of 20 kJ chemical energy, that is, the conversion factor being  $1 \text{ L/min} = 333 \text{ J/s} = 333 \text{ W}$ . Since also the

**TABLE 15.3**  
**Relative VO<sub>2</sub> max or HR, MET and RPE in Different Categories of Physical Activity Intensities**

| Intensity of Activity | Physiological Response              | Metabolic Measure | Subjective Perception    | Description of Whole-Body Dynamic Physical Activity                                                         |
|-----------------------|-------------------------------------|-------------------|--------------------------|-------------------------------------------------------------------------------------------------------------|
|                       | Relative HR/VO <sub>2</sub> max (%) | MET               | RPE                      |                                                                                                             |
| High                  | >85 to 100                          | >9                | (C): >17<br>(C-R): >7    | Work/exercise that cannot be sustained for more than 10 min                                                 |
| Vigorous              | >60 to 85                           | 6 to <9           | (C): 14–16<br>(C-R): 5–6 | Work/exercise in which a conversation cannot be maintained uninterrupted and may last up to about 30 min    |
| Moderate              | >40 to 60                           | 3 to <6           | (C): 11–13<br>(C-R): 3–4 | Work/exercise that can be conducted whilst maintaining a conversation uninterrupted and may last 30–60 min  |
| Light                 | 20 to <40                           | 1.6 to <3         | (C): 8–10<br>(C-R): 1–2  | Work/exercise that does not cause noticeable change in breathing rate and can be sustained for at least 1 h |
| Sedentary             | 0 to <20                            | 1 to <1.6         | (C): 6–7<br>(C-R): 0     | Sitting or lying with little additional movement                                                            |

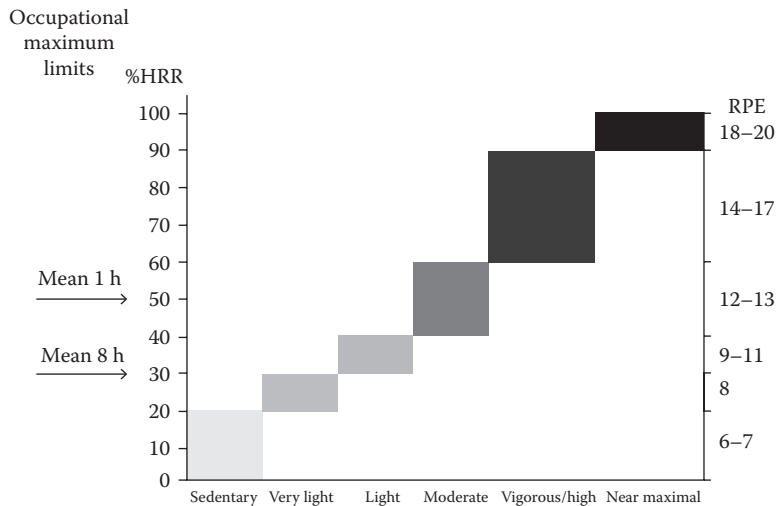
external mechanical work performed is measured in the Newtonian unit: J/s or Watt, this allows estimating the mechanical efficiency,  $E$ . The remaining chemical energy breakdown turns into heat. During dynamic work performed with large muscle groups,  $E$  is rather constant around 25% independent of work intensity but decreases with static work components. Consequently, 75% or more turns into heat.

$$\text{Mechanical efficiency} = \frac{\text{Energy produced as external mechanical work}}{\text{Total breakdown of internal chemically bound energy}}$$

Energy turnover at work can also be expressed as multiples of basal metabolic rate, which is called one metabolic equivalent task (1 MET unit), which roughly in oxygen uptake corresponds to 0.250 L/min. For instance, during walking, the energy expenditure may be about four times higher than that at rest. Then, the energy expenditure is about 4 MET units or 20 kJ/min (Table 15.3). That means that during walking, VO<sub>2</sub> is about 1.0 L/min or 14.3 mL/min/kg for a person weighting 70 kg. According to a position statement in 2012, physical activity intensity levels are divided into five categories (Table 15.3) (Norton et al., 2010). Of note is that these categories are composed mainly in relation to setting the framework for the recommendation for exercise training to conquer life style diseases due to physical inactivity. Such activity is performed in the order of 1/2–1 h per session up to once a day. In contrast, occupational physical activity has to be endured up to 8 h a day. This sets fundamentally different levels, which were presented for the first time as guidelines by the International Labor Organization in 1971, Bonjer, 1971 (Figure 15.4).

## BODY TEMPERATURE

Body temperature increases during muscle activity, since a large fraction of the chemical energy breakdown is converted to heat (approximately 75%) and only up to approximately 25% to mechanical work (i.e. the mechanical efficiency of work). To a certain extent, this is a regulated process, since core temperature rises within approximately 30 min to a constant level, which is higher, higher



**FIGURE 15.4** Relative HR and RPE for different intensity categories of work or physical activity combined with occupational maximum limits. (Reproduced from Sjøgaard, G. et al., *BMC Public Health*, 1, 652, 2014.)

the exercise intensity. That is, during prolonged exercise at 30% of  $\text{VO}_2$  max, body core temperature rises from 37.0°C to 38.0°C. With higher temperature, the chemical energy turnover processes run faster, and increased temperature in that sense is a benefit.

However, excessive heat must be dissipated because human tissues have a limited tolerance to high temperature. Likewise, low temperatures imply a risk factor, and core temperature should be maintained close to 37°C and skin temperature between 15°C and 40°C. The excessive heat produced during muscular activity is dissipated by radiation, convection and sweating. The former two are the main methods used during low temperatures, whereas sweating is the only possibility in high environmental temperatures. With an ambient temperature above 37°C, or in intense radiating heat, external heat is transferred to the body and must additionally be dissipated through sweating. Since sweating demands increased blood circulation to the skin, heat exposure causes an added demand on the circulatory system. Hence, muscle work in hot environment induces increments in HR far above those obtained at the same workload in a cold environment.

Prolonged heavy sweating also leads to fluid loss, which can severely reduce blood volume and thereby further increase the cardiovascular strain and which simultaneously leads to the loss of electrolytes crucial for tissue homeostasis. A loss of 1% of body mass through sweating leads to a deterioration of physical work capacity and reduced orthostatic tolerance. The World Health Organisation recommends that sweat production should not exceed 4 L in an 8 h work shift. Thus, heavy work in hot environment imposes a heat stress on the worker that may become excessive. This stress is further exaggerated in a humid environment where heat dissipation by sweating is hampered. Therefore, environmental measures of ambient temperature must be combined with measures of humidity and radiating heat to assess the heat stress of a worker, which must be taken into consideration relative to the exercise intensity. Direct predictions of heat stress can be obtained only through physiological measures of HR, core and skin temperatures, and fluid as well as electrolyte balance (see Chapter 24 for further discussion of management of and response to heat stress).

## ASSESSMENT OF WORK CAPACITY

Practical guidelines usually consider the measured external load such as the energy consumption of a given work task or the weight of a given burden. However, this is for the individual worker only a crude indication of the risk of overload since the relative load is highly dependent on the

relation between the external load and the actual capacity of the worker. Therefore, evaluation of the relative load constitutes a cornerstone within an ergonomic evaluation. For both the metabolic and mechanical aspects of work, the relative load can be calculated as the work demand in percentage of the relevant individual capacity. However, both the strength and the aerobic capacity of the individuals depend on a number of factors; among those, one of the essentials is the anthropometry.

## ANTHROPOMETRY

Detailed anthropometric measures of a given population are important not only for an estimation of the size-related range in strength and aerobic capacity but also for designing ergonomic workplaces and for biomechanical calculations used to quantify the workload. The range of anthropometric measures between small and large males and females indicates how flexible the workstations need to be to provide ergonomically suited work postures and movements for the major part of the working population. Anthropometric measures of human segment lengths are relatively simple using tape measures, slide gauges or inch rules. However, only few population studies provide such data. See Chapter 27 for further consideration of the use of anthropometry in work design.

The following is as an example given the results of an anthropometric survey of 467 Danish employees (231 females and 236 males). The study was performed in a mobile laboratory, visiting 84 cities in 9 counties in Denmark. The anthropometric measurements were standardised in relation to anatomic landmarks, and if unilateral, they were measured on the right side of the body. The following 14 anthropometric registrations were performed: weight, height, shoulder width, hand grip diameter, sitting height, as well as length of: the back, the entire leg, the lower leg, the entire arm, the upper arm, the forearm and hand, the hand, and the palm. Eleven of the anthropometric measurements in the study are illustrated in Figure 15.5 together with a detailed description of the anatomical landmarks used for definition. The results are for females and males shown in Table 15.4 (Faber et al., 2006 and Hansen et al., 1986).

Based on existing studies providing such detailed registration of body dimensions, the relation of all segment lengths to one simple measure of the body height can be estimated (Dempster, 1955, Plagenhoef, 1971, Phesant and Haslegrave, 2005). In Figure 15.6 is shown an overview giving the relative length of body segments that may be needed in an ergonomic evaluation of a workplace or a biomechanical calculation of the individual workload. As it can be seen in Table 15.4, the body height in the population almost ranges from 1.5 to 2.0 m, which implies that a workplace should be highly adjustable in order to offer optimal ergonomic conditions for all employees.

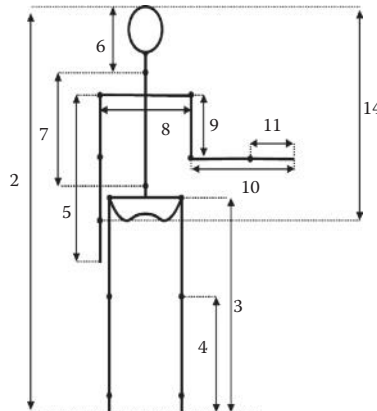
## RELATIVE MUSCLE LOAD

The maximal muscle force substantially depends on the amount of muscle mass, since individual muscles of same size and training produce the same amount of force independent of age and gender. However, the amount of muscle mass greatly differs with size, age and gender of the individual worker, and therefore force capacity shows a corresponding variation.

Muscle force depends on the physiological cross-sectional area of the muscle, which is related to the general size of the person. As an example, consider a difference in body height as shown in Table 15.4 and compare a 1.5 m tall person to a 2.0 m tall person. Everything else being equal, the difference in physiological cross-sectional area of all muscles solely due to the different body size of the persons will be as the second order of their height. Therefore, for the small person, the muscle cross section proportion will be  $1.5 * 1.5 = 2.25$  compared to the taller persons  $2 * 2 = 4$ ; in other words, the physiological cross-sectional area will be almost twice as large for the tallest person, and correspondingly maximal force will be almost twice as large compared to the smaller person (Figure 15.7a).



1. Weight
2. Stature
3. Leg length – floor to the lateral trochanter.
4. Lower leg length – floor to the lateral epicondyl of femur.
5. Arm length – upper edge of acromion to the tip of the third finger.
6. Head-neck length – vertex to processus prominens.
7. Spine length – processus prominens to midpoint between the crista iliace.
8. Shoulder width – right acromion to the left acromion.
9. Upper arm length – acromion to olecranon (elbow flexed 90°).
10. Lower arm and hand length – olecranon to the tip of the third finger. (elbow flexed 90°).
11. Hand length – proximal extension fissure at the wrist to the tip of the third finger.
12. Palm length of the hand – proximal flexion fissure of the wrist to the proximal flexion fissure at the third finger. (not shown)
13. Handgrip diameter – the circumference when the first and third finger just reach each other. (not shown)
14. Sitting height – sitting surface to the top of the head.



**FIGURE 15.5** Definition of anthropometric measures. (No 1, 12, and 13 are not indicated in figure.)

In general, a large gender difference exists in muscle force. For muscle groups in the upper extremities, men are almost twice as strong as women, while for the lower extremities, women have about 70% of the strength that men have. These differences are partly due to size differences, but even in men and women who are the same size, the muscle mass of a woman is on average about 80% of the muscle mass for a man (Figure 15.7b).

Regarding age, a decrease in muscle force is partly due to loss of muscle mass and partly due to a decreased ability of ageing muscle to generate tension, for example a change in properties of the muscle fibres. The decrease in muscle mass starts at about age 40, and on average, it is estimated that muscle force decreases about 1%–2% per year. This means that at an age of 65, muscle force may be only about 80% of lifetime best (Figure 15.7c). In particular, the ability to create force as fast as possible, muscle power decreases with about 3%–4% per year.

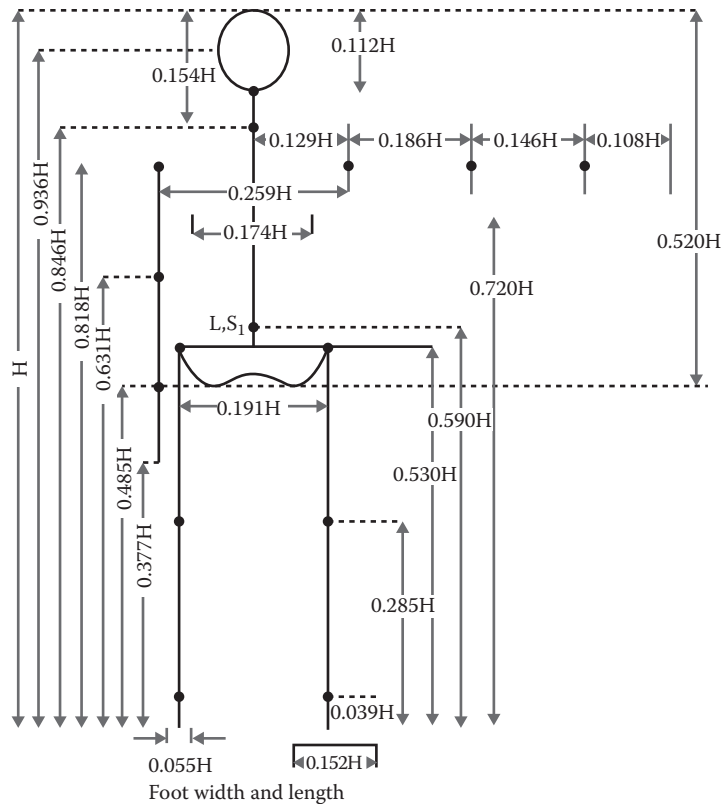
Taken together the differences due to size, gender and age, everything else being equal means that the muscle force capacity of a 2 m tall 24-year-old man may be about three times larger than the capacity a 1.5 m tall 65-year-old female as illustrated in Figure 15.7d. As an example, data giving the actual variation in muscle strength in relation to age and gender in a Danish population study are shown in Table 15.5. As it can be seen, the age-related difference is quite small in the occupationally active age groups, while the combined effect of size and gender is about a factor of two.

While size, gender and age are given factors that always must be considered in an individual ergonomic evaluation of a job task, it should be remembered that the premise is everything else is equal. But usually, this is not the case. A fourth factor, the fitness of the participant, may be the single most important issue since regular training over years easily can double the muscle strength. If training vs. no-training is applied to the extremes in the earlier example, the difference may actually change to a fivefold difference.

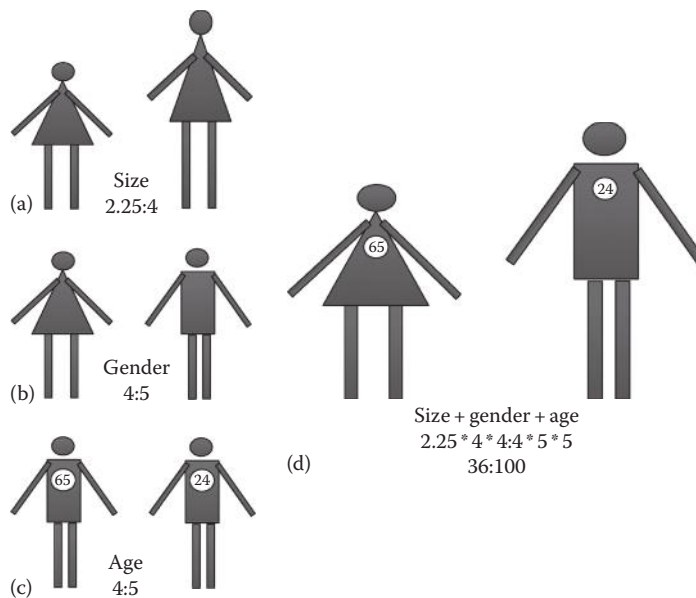
This means that an external load may represent a relative load of only 20% of maximal capacity for some employees but be almost close to maximum for others. Unfortunately, it is very difficult to apply regulations on an individual level based on relative load, and therefore guidelines are usually a compromise that place most workers within a safe range of exposure.

TABLE 15.4  
Results for 14 Anthropometrical Measures for 467 Danish Employees

|                              | Females (N = 231) |      |      |          |         |         | Males (N = 236) |      |      |          |         |         |
|------------------------------|-------------------|------|------|----------|---------|---------|-----------------|------|------|----------|---------|---------|
|                              | Fraction          |      |      | Fraction |         |         | Fraction        |      |      | Fraction |         |         |
|                              | Mean              | SD   | 5%   | 95%      | Minimum | Maximum | Mean            | SD   | 5%   | 95%      | Minimum | Maximum |
| Weight (kg)                  | 66.8              | 11.3 | 52.6 | 86.3     | 46.9    | 111.5   | 81.4            | 12.4 | 64.5 | 103.1    | 52.7    | 125     |
| Height (mm)                  | 1656              | 57   | 1560 | 1750     | 1525    | 1840    | 1774            | 69   | 1670 | 1883     | 1575    | 1960    |
| Leg length (mm)              | 844               | 43   | 773  | 915      | 740     | 950     | 907             | 44   | 835  | 975      | 775     | 1010    |
| Lower leg length (mm)        | 463               | 31   | 420  | 515      | 370     | 580     | 498             | 30   | 440  | 545      | 410     | 590     |
| Arm length (mm)              | 716               | 33   | 670  | 775      | 625     | 800     | 788             | 35   | 735  | 846      | 675     | 875     |
| Head-neck length (mm)        | 252               | 15   | 228  | 275      | 210     | 295     | 262             | 21   | 230  | 295      | 215     | 390     |
| Spine length (mm)            | 406               | 25   | 360  | 448      | 345     | 490     | 449             | 33   | 400  | 510      | 360     | 550     |
| Shoulder width (mm)          | 368               | 17   | 340  | 395      | 325     | 420     | 412             | 20   | 375  | 446      | 350     | 470     |
| Upper arm length (mm)        | 338               | 20   | 308  | 370      | 275     | 435     | 370             | 21   | 340  | 405      | 315     | 425     |
| Lower arm + hand length (mm) | 437               | 20   | 405  | 473      | 390     | 490     | 483             | 22   | 454  | 515      | 380     | 575     |
| Hand length (mm)             | 177               | 11   | 160  | 195      | 150     | 220     | 194             | 14   | 175  | 215      | 170     | 295     |
| Palm length (mm)             | 104               | 6    | 95   | 115      | 90      | 125     | 114             | 7    | 105  | 125      | 95      | 140     |
| Handgrip diameter (mm)       | 49                | 3    | 44   | 54       | 38      | 57      | 53              | 4    | 47   | 59       | 43      | 63      |
| Sitting height (mm)          | 876               | 29   | 830  | 925      | 775     | 970     | 924             | 38   | 860  | 986      | 800     | 1035    |



**FIGURE 15.6** Segment length calculated relative to height. (With reference to Sjøgaard, G., *Biomekanik*, in: *Basisbog i arbejdsfysiologi, bind I*, Arbejdsmiljøinstituttet, København, Denmark, 1995, pp. 66–94.)



**FIGURE 15.7** The range in muscle strength due to size (a), gender (b) and age (c) as well as the cumulated effect of size, gender and age (d).

**TABLE 15.5**  
**Muscle Strength in a Representative Study of Danish**  
**Female and Male Employees in Different Age Groups**

| Age (years)              | Females |       |     | Males |       |     |
|--------------------------|---------|-------|-----|-------|-------|-----|
|                          | <30     | 30–44 | >44 | <30   | 30–44 | >44 |
| Back extension (N m)     | 107     | 116   | 105 | 196   | 215   | 197 |
| Back flexion (N m)       | 99      | 98    | 95  | 186   | 194   | 180 |
| Arm abduction (N m)      | 33      | 32    | 31  | 75    | 70    | 67  |
| Shoulder elevation (N m) | 58      | 55    | 54  | 126   | 126   | 110 |
| Handgrip (N)             | 349     | 358   | 334 | 567   | 558   | 514 |

For many jobs in modern work life, maximal muscle force may not in itself be a limiting factor for the performance of a task once or a few times. However, the maximal muscle force defines the relative load for a given external burden, and this is almost linearly related to the endurance time. Therefore, another factor that must be included in an analysis is the duration or repetition of the work task. Depending on the relative load, muscle fatigue will develop during hours of performance of the same task. The worker with the highest relative load will also have the largest fatigue-related decrease in capacity that further increases the relative load. This may call for a regulation of the duration of a certain work task or the interruption with pauses to prevent fatigue development.

### RELATIVE METABOLIC LOAD

The earlier considerations regarding factors influencing strength capacity also apply for aerobic capacity. The difference due to size alone in maximal oxygen uptake between a 1.5 m and a 2.0 m tall person is again about a factor of two. The average oxygen uptake is for a 20-year-old person about 3.5 and 2.5 L/min for a man and a woman, respectively. Part of the difference is due to the average larger size of men, but a considerable gender difference exists even when calculating oxygen uptake per kg body weight. From the age of 20 years, the maximal oxygen uptake in general decreases with about 10% per 10 years of age for both men and women (Åstrand, 1960, Åstrand and Rodahl, 2002).

A work task can be classified in accordance with the oxygen cost it takes to do the work. If it is assumed that the work is performed with the same skill and coordination, then the cost is the same for all workers. For example, a job as cleaner is characterised as light to moderate work based on an average oxygen demand of 1 L/min. This means that for an average 55-year-old woman, the relative load of such a job would be higher than 50% of her capacity, and for most men, it would be higher than 33% of the maximal oxygen capacity.

Unlike the guidelines for lifted external burden, guidelines suggested for aerobic load during workday are based on the relative load of the worker. Maximal accepted relative load during a workday has been suggested at levels from 33% and up to 50% of maximal aerobic capacity. As mentioned earlier, a level of 33% would be considered as a moderate level if the task is sustained for only 30–60 min (Table 15.3). However, for an 8 h workday, the International Labor Organization and others have suggested 33% as the highest acceptable level and has a number of practical suggestions on reduction in work time if the task exceeds this demand. This means that an 8 h working day at 33% will be a physically active and demanding job but nevertheless not reach an intensity known from leisure time activity to have a training effect.

The general agreement of a level not exceeding 50% of maximal capacity is based on at least two elements. First, it has been showed that a freely chosen cadence in physically heavy work results in approx. 35%–40% of maximal aerobic capacity, and second, a workload above 50% results in increasing HR and increasing lactate level indicating that this is not a steady-state condition.

The existing guidelines for physically heavy work are based on the relative workload. While the oxygen cost of a given work task is nearly constant, then the capacity of the worker is depending on size, gender and age. Therefore, the work organisation must be highly flexible and adjustable to fit the task to the actual work capacity of the worker in order to stay within the limits of an acceptable relative workload. This is the case both in a given work situation adjusting to different individuals as well as in a working life perspective of the same individual since the decreasing capacity with age means an increasing relative workload of the same task (Ilmarinen, 1992).

Aerobic capacity is highly dependent on the training status of the person. Thus, a difference in fitness status can increase the range of the extremes in individual capacity. However, the other side of the coin is that the capacity of a worker is not a given factor but is also modifiable; improving this presents a tool for the general health protection of the worker.

### WORK REQUIREMENT/CAPACITY RATIO: HEALTH OUTCOMES

A model presenting the relation between the exposure, the acute response, the long-term effect and how the individual capacity interacts with these relations is shown in Figure 15.8. It is important to notice that the model is a rather crude attempt to understand how a work-related disorder such as musculoskeletal pain may develop. However, due to the multifactorial nature of musculoskeletal disorders, the model cannot include all risk factors and their interaction, and certainly musculoskeletal disorders also may develop due to other factors that are not work related. A number of various models have been presented over the years – one of the most frequently cited and including the possible mechanisms is the conceptual model for work-related musculoskeletal disorders published in 1993 in the *Scandinavian Journal of Work Environment and Health* (Armstrong et al., 1993).

The exposure represents the absolute external work demands that can be specified, no matter who performs the work, and it may contain the total work situation. Not only size, shape and weight of handled objects are important parts of the description but also the work tools and workstation are part of the exposure since they in many situations define the possible work posture. Using the example of a cleaner, the description of the exposure may consist of the work task floor cleaning, specified by the square metre of floor to be cleaned, the tool that is provided as well as the time allowed for the task and the requested quality of the cleaning. It can also be described as a work task requiring an oxygen consumption of about 1 L/min.

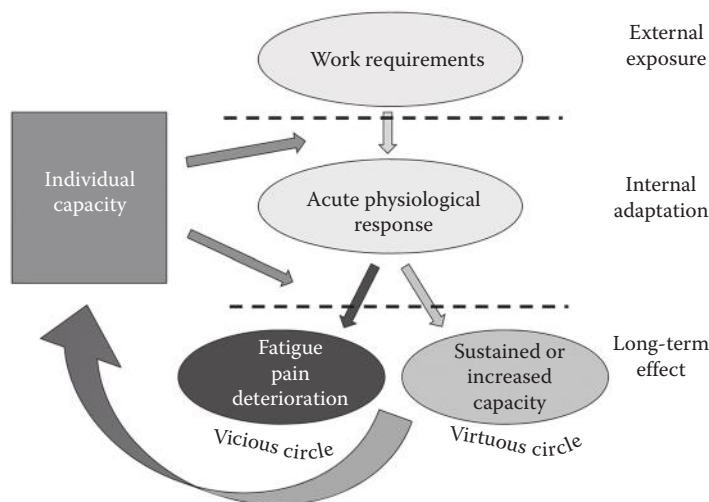


FIGURE 15.8 Exposure–response–effect model.

For any person who performs the requested external work, an internal adaptation will take place as a muscular mechanical action in muscles, tendons and ligaments and a corresponding metabolic change in order to deliver the needed fuel for the energy conversion to exerted force. This step in the model covers a cascade of physiological events and is termed acute physiological response.

If the work task is perpetually continued, eventually, it will result in the overload of the involved tissue, causing fatigue, pain, exhaustion and deterioration of the individual capacity. In contrast, if sustained or repeated for only prescribed intensive intervals and interrupted by rest periods allowing for restitution, the task may induce a training effect and increase or maintain capacity. These two different scenarios are both covered by the long-term effect in the lower part of the model (Figure 15.8) (Holtermann et al., 2007 and Holtermann et al., 2012).

As already pointed out in this chapter, the capacity of the actual worker can differ considerably, and this will be decisive for the level of the acute response and the relative load in the different steps of the evoked cascade of physiological adaptation to the work requirements.

The individual capacity may be acutely as well as in a long-term perspective reduced or improved due to previous exposure. Therefore, the individual capacity potentially will influence the relation both between external work and acute response as well as between acute response and the long-term response and the need of recovery. This can in a working life perspective be seen as a vicious circle where the accumulated long-term effect gradually decreases the capacity, causing an increase in the relative load and acute response of a given external work task and thereby enhances the negative long-term effect. However, the opposite situation of an initiated virtuous circle is also possible, initiating a training effect that increases individual capacity, thereby lowering the acute response and prolonging work endurance.

## PERCEIVED WORKLOAD

There is a curvilinear relationship between the intensity of a range of physical stimuli and the perception of their intensity. A positively accelerating relationship has been found between physical workload and perceived exertion in psychophysical studies. Thus, there is a highly reproducible individual relation between, for example, workload on a cycle-ergometer and perceived exertion. In the rating of perceived exertion (RPE) scale, the scale steps have been adjusted so that the ratings from 6 to 20 are linearly related to HR divided by 10 (Figure 15.9). The scale is presented to the test person before the start of the exercise test, and the endpoints of the scale (6 and 20) are thoroughly defined by verbal explanations. The scale is then shown to the test person at the end of each exercise load, and he or she is asked to rate the perceived exertion by giving the relevant number. The verbal explanations are used as support information. Also a nonlinear scale from 1 to 10 (CR10) with ratio properties is available for assessments in the laboratory or field (Figure 15.9). These scales are also termed *Borg scales* referencing the researcher first presenting and validating them (Borg, 1970 and Borg et al., 1985).

These scales of perceived exertion can be used to supplement physiological measurements during exercise and occupational work. They often provide valuable additional information about subjective responses, especially in cases where HR values are unreliable for some reason. For instance, HR responses to external muscle load may be *too weak* when major contributions of static work occur or *too strong* when work is performed with small muscle groups. Similar subjective scales have also been developed to quantify the intensity of pain.

RPE scales can be used to evaluate perceived exertion in actual work situations and give a fair estimate of the physiological load. The idea of substituting physiological measurements by perceived evaluations is attractive, as ratings do not require any instrumentation. However, findings in industry and in industrial tasks show that ratings are influenced not only by the overall perception of exertion but also by previous experiences and motivation of the workers. Thus, highly motivated workers may tend to underestimate their exertion and their physical workload (see Chapter 4 for further discussion of issues associated with subjective ratings).



| Scales for RPE                                     |                                                     |
|----------------------------------------------------|-----------------------------------------------------|
| Whole-body rating<br>perceived cardiovascular load | Local body region rating<br>perceived muscular load |
| 6                                                  | 0 Nothing at all                                    |
| 7 Very, very light                                 | 0.5 Very, very light                                |
| 8                                                  | 1 Very light                                        |
| 9 Very light                                       | 2 Fairly light                                      |
| 10                                                 | 3 Moderate                                          |
| 11 Fairly light                                    | 4 Somewhat hard                                     |
| 12                                                 | 5 Hard                                              |
| 13 Somewhat hard                                   | 6                                                   |
| 14                                                 | 7 Very hard                                         |
| 15 HARD                                            | 8                                                   |
| 16                                                 | 9                                                   |
| 17 Very hard                                       | 10 Very very hard (maximal)                         |
| 18                                                 |                                                     |
| 19 Very, very hard                                 |                                                     |
| 20                                                 |                                                     |

**FIGURE 15.9** Scales for the rating of perceived exertion. (Borg, 1970 and Borg et al., 1985).

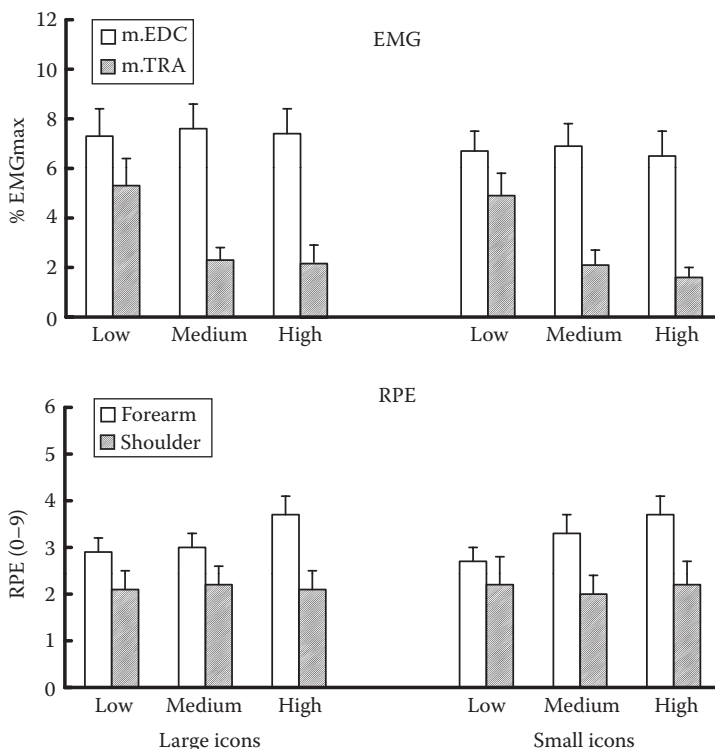
Importantly, although the relative workload can efficiently be assessed by subjective rating, this method has limited reliability compared to physiological measures. Therefore, physiological measures for workload assessments are requested. Examples of the use of RPE are given later in text showing levels that are attained during occupation work and possible pitfalls.

### MUSCLE WORK AND PERCEIVED LOAD DURING MONOTONOUS WORK

Computer mouse work requires highly intensive hand–eye coordination and has been studied extensively by electromyography (EMG, see earlier) and by RPE. The response patterns from these two methods are not always in concert. The following standardised case will highlight some points where results from the two methods demonstrate controversies, that is, the perceived load may underestimate the actual physiological muscle loading.

#### Computer Mouse Work Case

Twelve healthy right-handed females and males with a mean age of 35 years, without medical history of shoulder or neck problems, volunteered in this study. They performed a standardised computer task with repeated mouse clicking for 5 min at free work pace at two different circular-shaped icon sizes: small (2 mm) and large (20 mm) on the screen. The mouse motion settings (Logitech MouseWare® 9.1 Windows) were set at three levels: medium (standard speed + acceleration setting), low and high (i.e. lowest and highest possible motion settings, respectively). The study design was a complete  $2 \times 3$  factor experiment with icon sizes: small and large, and mouse motion: low, medium and high. Productivity was assessed in terms of the number of graphs completed during the 5 min work period as well as the number of mouse clicks performed. RPE was assessed according to a 10-point scale for the forearm and shoulders, additional to an overall RPE. Surface EMG was recorded on the right side from the shoulder muscle: m. trapezius (TRA) and from the forearm muscle: m. extensor digitorum communis. EMG was analysed as amplitude in percentage of maximal EMG amplitude during an MVC. It was hypothesised that optimum mouse motion in terms of minimum muscle load depends on icon size and that perceived exertion is a valid measure.



**FIGURE 15.10** Mouse work EMG vs. RPE.

More specifically, we expected that with smaller icon size (i.e. increased precision demand), lower mouse motion is optimal to minimise EMG and maximise production rate. We assumed that RPE and EMG would show similar responses. The finding showed that productivity for the larger icons was about twice the value of the smaller icons, while mouse motion level had only minor effect (<10%). The overall RPE was highest for the small icons (3.9 vs. 2.9,  $p < 0.05$ ), while RPE for the forearm was highest for the highest level of mouse motion (3.7 for high vs. 3.2 for medium and 2.7 for low,  $p < 0.05$ ), but RPE for the shoulder was independent of icon and mouse settings and in the order of 2 on the 10-point scale (Figure 15.10).

Interestingly, EMG was dissociated from RPE: for the shoulder muscle, EMG was significantly higher for the low mouse motion (5.1%) compared to medium (2.2%) and high (1.9%) mouse motion, while for the forearm muscle, only the icon size had a significant effect (6.7% for small and 7.4% for large icons), which may be due to the higher production rate with large icons (Figure 15.10). Increased mouse motion did not affect productivity but increased forearm and total RPE, in spite of decreased EMG in the shoulder muscle. It was surprising that RPE for shoulder and forearm did not vary in line with the respective EMG measures. Also the relative exertions were at quite different levels: EMG was approx. 2%–7% of its maximum, while RPE was in the order of 20%–40% of the maximum of the RPE scale used. This is probably due to the perceived exertion also including visual and coordination demands. The findings emphasise that sensory feedback by biophysical measures (RPE) during low-level muscle activity poorly reflects the actual muscle activity level.

### PERCEIVED LOAD IN PHYSICALLY HEAVY WORK

In jobs with whole-body physical loading, usually the scale 6–20 is used. In a variety of physically heavy jobs, these measures have reported values around 13: industrial workers 13.2 (10.1–15.2), construction workers 11.1 (9.5–12.8) and health care workers 13.1 (9.1–15.1). For cleaners, several

studies have similarly shown that the perceived load during floor cleaning on average is scored around 13. Remarkably, if the cleaners are asked over time, it has been shown that at the start of work, the load is scored only around 6, but within 1 h of floor cleaning, it reaches 13. Interestingly, these examples of workers in heavy jobs as an overall means assess their workload to be ~40% of their maximal capacity, which is quite high and above maximum level according to HR values for an 8 h workday. Important are also the large ranges implying that some are relatively much higher exposed than others. In summary, reports on perceived exertion give valuable complementary information on the subjective load during the job; however, they cannot substitute objective biomechanical and physiological measures assessing the load during different job tasks for risk assessment (Takala et al., 2010).

## CONTEMPORARY PERSPECTIVES IN ERGONOMICS/HUMAN FACTORS

In the classical understanding of the discipline of E/HF, it has for many decades been defined as *fitting the task to the person*. This means primarily adjusting the workplace to provide the most optimal work condition for as large a range of workers as possible. Since both dimensions and capacities of workers present with considerable ranges and may change over time, the workplaces should be highly adjustable. Conversely, during the recent decade, an emerging understanding of the importance of also to some degree *fitting the person to the task* has developed. Among other factors, this is due to the focus on the general positive effect of physical exercise and the great increase in inactive computer users due to the global development of information technology. This has caused a major concern for the adverse health effect of sedentary work hours and lifestyle in general.

Traditional occupational jobs that are well defined in time and location may still primarily call for E/HF concerns, for example, for limits of the physical exposure at the specific workstation and adjustment of the task with a focus on the relative physiological load on the worker. However, for a large part of the workforce, the work demands have changed so the maximal capacity and the relative workload are of minor importance, and it may even be the lack of activity and variation in work demands that form the largest risk factor for sustained health. Furthermore, among computer workers, the portable IT technology enables work to be performed from anywhere and at flexible times of the day. This calls for an E/HF approach that follows the worker and focuses on improving the workers' capacity which in a long-term perspective may benefit both a sustained work ability and health in general (Kadefors et al. 2002).

In a public health perspective, much focus has been on the decrease in occupational activity over the last century and especially the increase in sedentary work in the last decades. On the other hand, it has for many years been known that work-related activity does not have the same positive effect on the maintenance of physical capacity as leisure time activity or planned physical exercise. For workers in physically heavy jobs and no leisure time activity, a considerably increased risk of cardiovascular death has been demonstrated. However, while physically heavy work greatly increases risk for workers with low aerobic capacity, no increased risk was present for the physically fit workers. Therefore, also for the worker, in a physically heavy job, it can be crucial to add leisure time physical activity in order to preserve strength and aerobic capacity and thereby maintain work ability.

## COMBINING IMPROVED ERGONOMICS AND TAILORED PHYSICAL EXERCISE

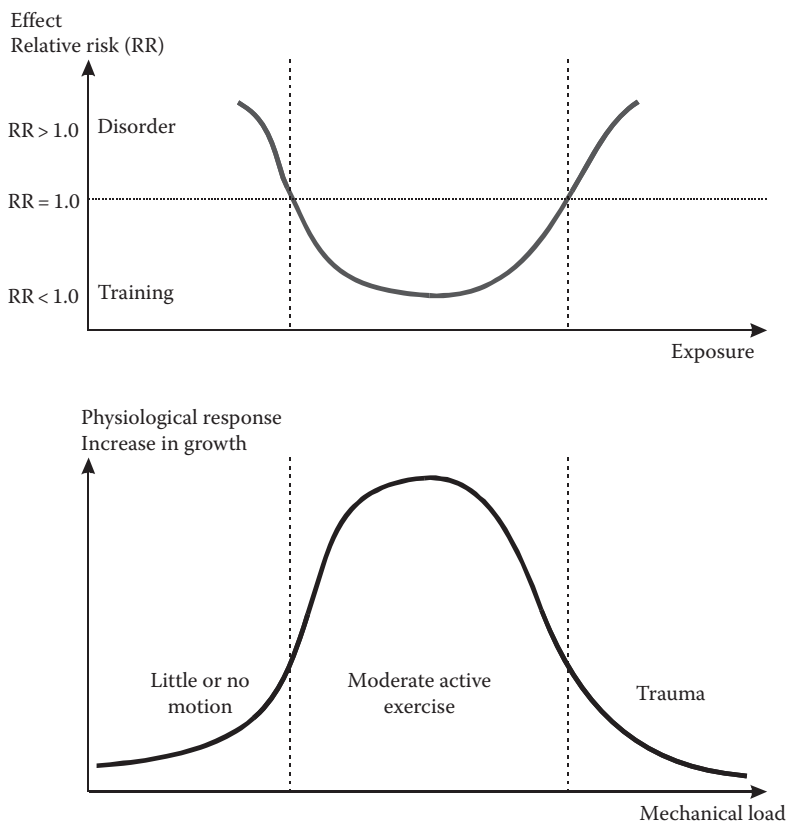
To summarise the principles for optimising work requirement, these involve minimising peak loads, limiting time with high mean force, maintaining dynamic work activities and allowing for as large a variation in physical load as possible (Table 15.6). If the possibilities for variation of the work task are limited, this may be introduced by providing possibility for physical exercise training to maintain work capacity. Even in work tasks classified as physically heavy work, the task may lack the timing and intensity needed to incite a training effect, and this needs to be approached by tailored physical exercise training. As a consequence, sedentary work and physically heavy work may call for

**TABLE 15.6**  
**Categories of Overload in Work Exposure**

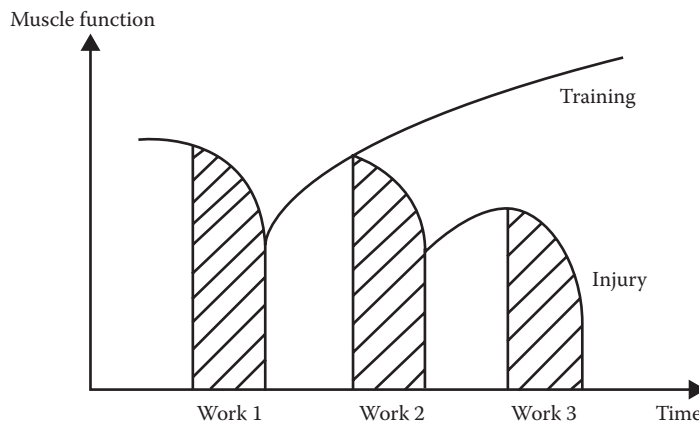
- Peak  
Manual material handling, heavy lifting, unexpected loads  
(in all occupations, e.g. accidents)
- Mean  
Physically heavy work with high energy consumption  
(in construction work, agriculture, fishery, cleaning)
- Accumulated  
Monotonous, highly repetitive work  
(in the meat industry, assembly line work, textile industry)

a totally different type of physical exercise training in order to provide a physiologically beneficial effect on the worker. Therefore, a thorough analysis of the work exposure as well as the capacity of the workers may form the basis for a sound exercise profile tailored to different occupational sectors and combining optimal E/HF adjustment of the working conditions with physical exercise training.

These principles may apply both in sedentary and in physically heavy jobs. The relation between the exposure and the effect is not a simple linear relation indicating that the more exposure



**FIGURE 15.11** Optimising exposure through intensity of mechanical load. (Reproduced from Sjøgaard, G. and Jensen, B.R., Muscle pathology with overuse, in: *Chronic Musculoskeletal Injuries in the Workplace*, Ranney, D. (ed.), W.B. Saunders Company, Philadelphia, PA, 1997, pp. 17–40.)



**FIGURE 15.12** Effect optimisation through duration of exposure and rest periods. (Reproduced from Sjøgaard, G. and Jensen, B.R., Muscle pathology with overuse, in: *Chronic Musculoskeletal Injuries in the Workplace*, Ranney, D. (ed.), W.B. Saunders Company, Philadelphia, PA, 1997, pp. 17–40.)

the worse. Rather, this relation forms a U-shape with an optimum exposure for obtaining the largest physiological benefit in terms of tissue growth, strength and general function (Figure 15.11). This also implies that in each end of the spectrum, underload as well as overload presents a non-optimal situation or a risk factor for deterioration of the tissue. In a time perspective, an accumulated loading without the adequate recovery could lead to a gradual decrease in capacity and so could accumulated inactivity. In contrast, the right match of the intensity of activity and suitable recovery period would induce an increased growth and training effect (Figure 15.12). This clearly demonstrates the contrasting effects of muscle activity on health, and these contrasting effects depend not only on the mode of muscle activity but also on the intensity, duration and repetition of the activity (Holtermann et al., 2012).

## PHYSIOLOGICAL WORK ASSESSMENT AS A PREMISE FOR A SUSTAINED WORKFORCE

The large variety of occupations and difference in tasks within each occupation underlines that the physical activity at the job will vary tremendously between workers. Add to this that the capacity of workers covers a wide range when taking anthropometry, age and gender into account. This results in dramatic differences when it comes to relative physical loading of the body at work. Workers may be underexposed to physical activity from a health perspective, simply due to inactivity at work and lack of time for exercise training at leisure. The fact is that duration and intensity of physical activity are crucial for health benefits, and these benefits may not ever be met for all employees, for example, in industries that routinely expose workers to physical heavy work.

The international recommendations for health are that all healthy adults aged 18–65 years should as a minimum engage in moderate-intensity cardiorespiratory exercise for more than 30 min/day on  $\geq 5$  day/week for a total of  $\geq 150$  min/week and vigorous-intensity cardio respiratory exercise training for  $\geq 20$  min/day on  $\geq 3$  day/week ( $\geq 75$  min/week) or use a combination of moderate- and vigorous-intensity exercises to achieve a total energy expenditure of  $\geq 500$ – $1000$  MET min/week. Additionally, it is recommended to perform resistance exercise 2–3 day/week for the major muscle groups and/or for those muscles causing pain. These target levels for physical activity in terms of duration and intensity can be obtained by accumulating smaller periods of physical activity, for example, by splitting 30 min in three bouts lasting 10 min or more (Thompson et al., 2003 and World Health Organization, 2010).

Monitoring physical activity during different jobs and tasks reveals that over a week, these recommendations are rarely fulfilled. Even among construction workers, this may not be the case and could account for low fitness in terms of oxygen uptake reported in this occupational group. Comprehensive programmes for improving employees' cardiovascular and general health have been recommended (2009, the American Heart Association). These programmes should include regular physical activity at moderate–vigorous intensity engaging large muscle groups and should encourage to healthy behaviour at work and leisure.

In the twenty-first century, an increasing number of randomised controlled trials, RCT, have been conducted introducing physical exercise training programmes at the worksite often performed during working hours. Growing evidence is that such programmes result in clinically relevant health effects, and preliminary cost-effectiveness estimates indicate acceptable cost relative to societal savings on health expenses. Worksite health promotion and job exposure–dependent exercise training are thus contemporarily recognised as a significant tool for health improvement in the workforce and thereby also of benefit to industry. Complementing work exposure profile to develop a healthy exposure/training profile integrating leisure time activity in work life is increasingly considered cost-effective. Examples of such approach are given later.

Nine RCT studies have recently been conducted in Denmark using the same basic intervention of offering supervised physical exercise training at the worksite during working hours according to the concept *intelligent physical exercise training* (Holtermann et al., 2010 and Sjøgaard et al., 2014). A wide variety of job categories with highly different physical activities were studied (Table 15.7). The generic concept implied the following three essential factors, which made them distinct from a number of unsuccessful interventions:

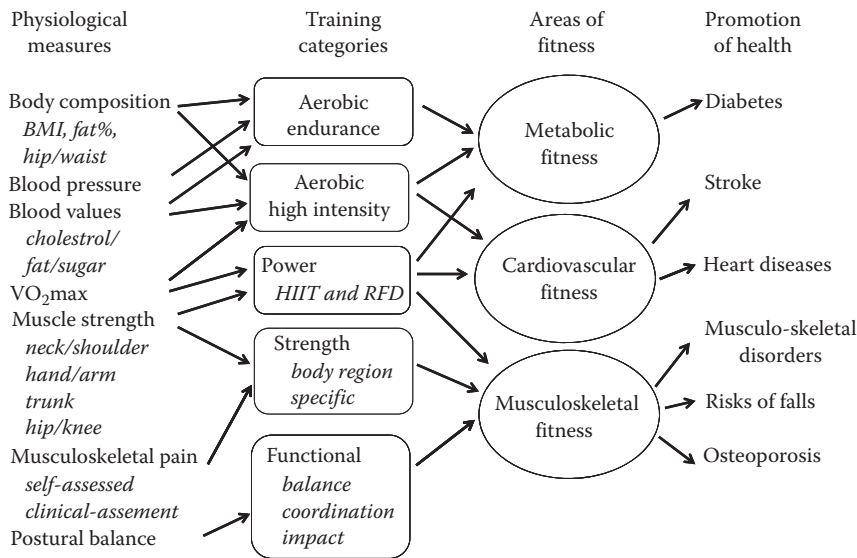
1. Physical exercise training was performed during working hours for 1 h/week, usually divided into two to three training sessions or broken down into, for example, periods of 10 min duration. This requested the involvement of the employer to allow for such activities, thus signalling the support of health enhancement for employees.
2. Sports exercise training specialists were involved in designing the specific exercise training programmes that were based on evidenced sports sciences training principles, of general high intensity, and tailored to work exposure, employee health status and physical capacity (based on a health check).
3. Training sessions were carefully introduced and supervised by expert trainees at the company level, and adherence was monitored for motivational purposes.

Figure 15.13 illustrates health check variables and how they relate to specific training recommendations based on evidenced sport and medical science training principles. Also, the areas of

**TABLE 15.7**  
**References for Nine Danish RCT Studies on Worksite Physical Exercise Intervention among Seven Different Job Groups**

| Job Category           | References                                                                                       |
|------------------------|--------------------------------------------------------------------------------------------------|
| Office workers         | <i>Scand J Work Environ Health</i> 34: 55–65, 2008<br><i>Br J Sports Med</i> 46: 1004–1010, 2012 |
| Computer workers       | <i>Arthritis Rheum</i> 59: 84–91, 2008<br><i>J Occup Rehabil</i> 24: 316–324, 2014               |
| Industrial technicians | <i>BMC Musculoskelet Disord</i> 12: 205, 2011                                                    |
| Cleaning personnel     | <i>BMC Public Health</i> 11: 776, 2011                                                           |
| Health care workers    | <i>BMC Public Health</i> 11: 671, 2011                                                           |
| Construction workers   | <i>Scand J Work Environ Health</i> 38: 467–475, 2012                                             |
| Fighter pilots         | <i>Clin J Pain</i> 29(1): 50–59, 2013                                                            |





**FIGURE 15.13** Exercise for health: physiological assessment as premise for a sustained healthy workforce. (Reproduced from Sjøgaard, G. et al., *BMC Public Health*, 14, 652, 2014.)

fitness and health are illustrated specifying areas for the prevention and rehabilitation of diseases. The health check may include a selection of variable among the earlier provided depending on the occupational challenge and the worker's health profile. Additional relevant variables may of course be added.

The interventions enrolled ~2500 workers within a wide range of job groups: office workers, computer workers, industrial laboratory technicians, cleaning personnel, health care workers, construction workers and fighter pilots. After a duration between 3 months and 1 year, questionnaire surveys and health checks showed that in all job groups, significant improvements were documented regarding health outcomes and/or health risk indicators. These outcomes were job group specific:

- Neck pain was reduced among office and computer workers, industrial laboratory technicians, cleaners as well as fighter pilots, and forearm pain was reduced among office/computer workers and laboratory technicians. Also decreases in back pain were reported. The magnitude of pain decreases was clinically relevant of 1 or more on 10-point rating scale (similar to the RPE scale) and decreased by 50% among those with high pain intensity, that is, above 3 on this scale.
- Aerobic capacity – a health risk indicator for cardio-metabolic diseases – was improved among office/computer workers, health care workers and construction workers. Workplace physical exercise training of just 1 h a week for 12 weeks induced a 10% increase in maximal aerobic capacity, for example, counteract an age-related 10-year decrease. Similarly, an increase in the functional aerobic capacity calculated as aerobic power per kg body weight significantly improved by a weight reduction among overweight workers.

In addition, a number of other improvements in physical capacities were evidenced such as increased muscle strength and balance control. It is important to emphasise that in every study group, outcomes of improved health were documented, and the effect sizes were of clinical relevance. It is concluded that worksite exercise training does enhance health if a programme with evidenced efficacy is implemented by expert trainees and supported by the employer.

## REFERENCES

- Aminoff T, Smolander J, Korhonen O and Louhevaara V. Prediction of acceptable physical work loads based on responses to prolonged arm and leg exercise. *Ergonomics*, 41: 109–120, 1998.
- Andersen LL, Kjaer M, Sjøgaard K, Hansen L, Kryger A and Sjøgaard G. Effect of two contrasting types of physical exercises on chronic neck muscle pain. *Arthritis Rheum* 59: 84–91, 2008.
- Armstrong TJ, Buckle P, Fine LJ, Hagberg M, Jonsson B, Kilbom Å, Kuorinka IAA, Silverstein BA, Sjøgaard G and Viikari-Juntura ERA. A conceptual model for work-related neck and upper-limb musculoskeletal disorders. *Scand J Work Environ Health* 19: 73–84, 1993.
- Armstrong TJ, Fine LJ, Goldstein SA, Lifshitz YR and Silverstein BA. Ergonomics considerations in hand and wrist tendinitis. *J Hand Surg Am* 12(5 Pt 2): 830–837, September 1987.
- Åstrand I. Aerobic work capacity of men and women with special reference to age. *Acta Physiologica Scandinavica* 49(Suppl. 169): 1–92, 1960.
- Åstrand PO. and Rodahl K. *Textbook of Work Physiology*. New York: McGraw-Hill, 1986; Wu HC and Wang MJJ. Relationship between maximum acceptable work time and physical workload. *Ergonomics*, 45: 280–289, 2002.
- Blangsted AK, Sjøgaard K, Hansen EA, Hannerz H and Sjøgaard G. One-year randomized controlled trial with different physical-activity programs to reduce musculoskeletal symptoms in the neck and shoulders among office workers. *Scand J Work Environ Health* 34: 55–65, 2008.
- Bonjer FH. Energy expenditure. In: *Encyclopedia of Occupational Health and Safety*, Parmeggiani L (ed.). Geneva, Switzerland: International Labour Organisation, pp. 458–460, 1971.
- Borg G. Perceived exertion as an indicator of somatic stress. *Scand J Rehabil Med* 2: 92–98, 1970.
- Borg G, Ljunggren G and Cegi R. The increase of perceived exertion, aches and pain in legs, heart rate and blood lactate during exercise on a bicycle ergometre. *Euro J Appl Physiol Occup Physiol* 54: 343–349, 1985.
- Christensen JR, Faber A, Ekner D, Overgaard K, Holtermann A and Sjøgaard K. Diet, physical exercise and cognitive behavioural training as a combined workplace based intervention to reduce body weight and increase physical capacity in health care workers – a randomised controlled trial. *BMC Public Health* 11: 671, 2011.
- Dempster WT. Space requirements of the seated operator. Geometrical, kinematic, and mechanical aspects of the body with special reference to the limbs. Technical report WADC-TR-55-159. OH: Wright-Patterson Air Force Base, 1955.
- Faber A, Hansen K and Christensen H. Muscle strength and aerobic capacity in a representative sample of employees with and without repetitive monotonous work. *Int Arch Occup Environ Health* 79(1): 33–41, January 2006.
- Garber CE, Blissmer B, Deschenes MR, Franklin BA, LaMonte MJ, Lee IM, Nieman DC and Swain DP. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Med Sci Sports Exerc* 43(7): 1334–1359, 2011.
- Gram B, Holtermann A, Sjøgaard K and Sjøgaard G. Effect of individualised work site exercise training on aerobic capacity and muscle strength among construction workers. A randomised controlled intervention study. *Scand J Work Environ Health* 38(5): 467–475, 1 September 2012.
- Hansen K, Feveile H, Finsen L, Jensen C, Jensen A, Merkel J, Ekner D and Christensen H. Kropsmål for danske arbejdstagere Antropometriske mål på 467 danske arbejdstagere. Arbejdsmiljøinstituttet, København, Denmark, 1998.
- Holtermann A, Hansen JV, Burr H, Sogaard K and Sjøgaard G. The health paradox of occupational and leisure-time physical activity. *Br J Sports Med* 46: 291–295, 2012.
- Holtermann A, Jørgensen MB, Gram B, Christensen CR, Faber A, Overgaard K, Ektor-Andersen J, Mortensen OS, Sjøgaard G and Sjøgaard K. Worksite interventions for preventing physical deterioration among employees in job-groups with high physical work demands: Background, design and conceptual model of FINALE. *BMC Public Health* 10: 120, 2010 (doi:10.1186/1471-2458-10-120).
- Holtermann A, Mortensen OS, Burr H, Sjøgaard K, Gyntelberg F and Suadicani P. Physical work demands and physical fitness in low social classes – 30-year ischaemic heart disease and all-cause mortality in The Copenhagen Male Study. *J Occup Environ Med* 53(11): 1221–1227, November 2007.
- Ilmarinen, J., Job design for the aged with regard to the decline in their maximal aerobic capacity: Part I-guidelines for the practitioner. *Int J Ind Ergon* 10: 53–63, 1992.

- Jørgensen MB, Ektor-Andersen J, Sjøgaard G, Holtermann A and Sjøgaard K. A randomised controlled trial among cleaners – effects on strength, balance and kinesiophobia. *BMC Public Health* 11: 776, 10 October 2011.
- Kadefors R, Christensen H, Sjøgaard G, Colombo R, Merletti R, Vallbo A, Lundberg U et al. PROCID recommendations for healthier computer work – recommendations. *Work Stress* 16: 91–93, 2002.
- Morris JN, Heady JA, Raffle PAB, Roberts CG and Parks JW. Coronary heart-disease and physical activity of work. *Lancet* 2(6795): 1053–1057, 21 November 1953.
- Norton K, Norton L and Sadgrove D. Position statement on physical activity and exercise intensity terminology. *J Sci Med Sport* 13(5): 496–502, September 2010.
- Panel on Musculoskeletal Disorders and Workplace. National Research Council and the Institute of Medicine. Musculoskeletal disorders and the workplace: Low back and upper extremities. Washington, DC: National Academy Press, 2001.
- Pheasant S and Haslegrave CM. *Bodyspace: Anthropometry, Ergonomics, and the Design of Work*. London, U.K.: Taylor & Francis Group Ltd., 2005.
- Plagenhoef S. *Patterns of Human Movement*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1971.
- Silverstein BA, Fine LJ and Armstrong TJ. Hand wrist cumulative trauma disorders in industry. *Br J Ind Med* 43(11): 779–784, November 1986.
- Sjøgaard, G. Biomekanik. In *Basisbog i arbejdsfysiologi, bind I*. København, Denmark: Arbejdsmiljøinstituttet, pp. 66–94, 1995.
- Sjøgaard, G. Muscles. In: *ILO Encyclopaedia of Occupational Health and Safety*, Riihimäki H and Viikari-Juntura E (eds.). Geneva, Switzerland: International Labor Office, pp. 6.2–6.4, 1997.
- Sjøgaard G. and Jensen BR. Muscle pathology with overuse. In: *Chronic Musculoskeletal Injuries in the Workplace*, Ranney D (ed.). Philadelphia, PA: W.B. Saunders Company, pp. 17–40, 1997.
- Sjøgaard G and Jensen BR. Low-level static exertions. In: *Fundamentals and Assessment Tools for Occupational Ergonomics*, Marras WS and Karwowski W (eds.). Boca Raton, FL: Taylor & Francis Group/CRC Press, 2006.
- Sjøgaard G, Jensen BR, Hargens A and Sjøgaard K. Intramuscular pressure and EMG relate during static contractions but dissociate with movement and fatigue. *J Appl Physiol* 96: 1522–1529, 2004.
- Sjøgaard G, Justesen JB, Murray M, Dalager T and Sjøgaard K. A conceptual model for worksite intelligent physical exercise training – IPET – intervention for decreasing life style health risk indicators among employees: A randomised controlled trial. *BMC Public Health* 14: 652, 2014.
- Sjøgaard G, Lundberg U and Kadefors R. The role of muscle activity and mental load in the development of pain and degenerative processes at the muscle cell level during computer work. *Eur J Appl Physiol* 83: 99–105, 2000.
- Sjøgaard G and Sjøgaard K. Muscle injury in repetitive motion disorders. *Clin Orthop* 351: 21–31, 1998.
- Sjøgaard G and Sjøgaard K. Myalgia, chronic. In: *Encyclopedia of Exercise Medicine in Health and Disease*, Mooren FC (ed.). Berlin, Germany: Springer, 2012.
- Takala EP, Pehkonen I, Forsman M, Hansson GA, Mathiassen SE, Neumann WP, Sjøgaard G, Veiersted KB, Westgaard RH and Winkel J. Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scand J Work Environ Health* 36: 3–24, 2010.
- Thompson PD, Buchner D, Piña IL, Balady GJ, Williams MA, Marcus BH, Berra K et al. Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease. A statement from the council on clinical cardiology (Subcommittee on Exercise, Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity). *Circulation* 107: 3109–3116, 2003.
- World Health Organisation. *Global Recommendations on Physical Activity for Health*. Geneva, Switzerland: World Health Organisation, p. 58, 2010.



---

# 16 Assessment of Physical Job Demands

*Jonisha P. Pollard and Patrick G. Dempsey*

## CONTENTS

|                                                               |     |
|---------------------------------------------------------------|-----|
| Introduction.....                                             | 448 |
| Initiating the Assessment.....                                | 448 |
| Approaches to Workplace Assessment of Physical Demands.....   | 449 |
| Survey of Certified Professional Ergonomists .....            | 449 |
| Job Demand Assessment Toolkit.....                            | 450 |
| Interviews and Questionnaires .....                           | 450 |
| Direct Observation .....                                      | 450 |
| Temporal Characteristics .....                                | 450 |
| Video .....                                                   | 451 |
| Direct Measurement .....                                      | 452 |
| Assessment Tools and Methods .....                            | 452 |
| Discomfort.....                                               | 452 |
| Posture and Motion .....                                      | 454 |
| Goniometer.....                                               | 455 |
| Industrial Lumbar Motion Monitor .....                        | 456 |
| Optoelectronic Systems .....                                  | 457 |
| Accelerometry .....                                           | 457 |
| Inertial Measurement Units .....                              | 457 |
| Manual Materials Handling.....                                | 458 |
| Revised NIOSH Lifting Equation (United States) .....          | 458 |
| 1992 Manual Handling Assessment Charts (United Kingdom) ..... | 459 |
| Psychophysical Tables .....                                   | 461 |
| Physiological Assessment.....                                 | 462 |
| Biomechanical Assessment .....                                | 466 |
| Upper Limb Assessments .....                                  | 466 |
| RULA .....                                                    | 466 |
| ART Tool (United Kingdom) .....                               | 467 |
| Force and Exertions.....                                      | 472 |
| Pressure Sensors .....                                        | 474 |
| Force Sensors .....                                           | 475 |
| Electromyography .....                                        | 478 |
| Whole-Body Assessment .....                                   | 478 |
| Ovako Working Posture Analysis System .....                   | 478 |
| Rapid Entire Body Assessment .....                            | 479 |
| Conclusion .....                                              | 484 |
| Disclaimer .....                                              | 484 |
| References.....                                               | 484 |

## INTRODUCTION

The assessment of physical demands in the workplace is an integral component of ergonomics/human factors (E/HF) for many types of jobs. The physical interface with the workplace, the forces exerted by muscles and energy expended by the body during work can lead to discomfort, fatigue, injuries and absence due to sickness that decrease human performance. Job demands range from postural restrictions common in jobs as diverse as office work and mining to high biomechanical and physiological demands found in jobs requiring exertion of high forces or metabolic expenditures. Contact stresses from protective clothing, tools and equipment can be additional sources of discomfort and acute injuries.

Because this chapter is intended to be fairly broad in scope, the second section will provide an overview of the types of measurement (e.g. observational, direct) that can be selected when conducting a workplace assessment of a job or task. Commonly used methods will be described, and trade-offs among measurement types will be discussed. Recommendations for a basic data collection tool kit will be provided based on a survey of certified professional ergonomists (CPEs) in the United States.

We then cover specific assessment tools and methods and is organised to provide an overview of the range of methods available to the ergonomist. There is no single approach to effective workplace assessment. There are often technical reasons, constraints and influences that impact choices, and a number of these will be discussed throughout the chapter. The advantage of the diverse possibilities is that the individual skills and preferences of the analyst can be accommodated by a variety of available analytical tools (Dempsey and Leamon, 1999; Dempsey et al., 2005). It is common for ergonomists to have favourite analysis tools.

Assessing work demands often requires measurement and assessment of individual parameters of the task, tools and equipment, environment and human responses (e.g. discomfort, force, energy expenditure). Musculoskeletal disorder (MSD) prevention is often described in terms of assessing the primary risk factors of force, posture and repetition. The measurement of these individual risk factors in the workplace will be discussed because they are often used directly by the ergonomist or as input to an assessment tool that provides a form of scoring or ranking – such as the National Institute for Occupational Safety and Health (NIOSH) Lifting Equation (NIOSH, 1981; Waters et al., 1994) and the Rapid Upper Limb Assessment (RULA; McAtamney and Corlett, 1993). Many of these assessment methods appeal to the ergonomist who is often faced with prioritising jobs or tasks for improvement.

Two of the most commonly performed assessments of physical work focus on related classes of tasks involving manual materials handling (MMH) or upper-extremity-intensive work. These two classes of assessment will be discussed separately for this reason. There is also a wide range of methods in the literature for ergonomists who do not want to develop their own approach using measurements of individual risk factors.

## INITIATING THE ASSESSMENT

Although proactive design of tools, equipment and the work environment to maximise human performance is the ideal case, ergonomists often assess physical job demands based on reports of discomfort, injuries or productivity issues that result from mismatches between job demands and human capabilities and limitations. Worker reports of discomfort or pain, insurance data, regulatory data on reported injuries and illnesses, and medical or first-aid visits are all potential triggers for a more in-depth analysis. Especially in manufacturing and related types of paced work, productivity issues are referred to as 'bottle necks', and the productivity problem is often a symptom of less-than-ideal E/HF design. Overall, many facilities assess physical demands over time with a focus on resolving problems or reducing exposures deemed most critical. Ideally, improving the match between physical demands and the capabilities and limitations of the workforce should be a continuing process.



## APPROACHES TO WORKPLACE ASSESSMENT OF PHYSICAL DEMANDS

This section provides the ergonomist with an overview of basic guidelines for tools and approaches to measuring physical demands. These guidelines are broad enough to allow the reader to customise their approach to workplace measurement by choosing those items that are consistent with their financial constraints, technical background and the purpose of the analysis at hand. For example, biomechanical models (including those commercially available) require an understanding of the underlying principles of mechanics and assumptions in order to correctly apply the model (see Chapter 17 by Dempsey and Munro). An ergonomist without the requisite engineering background may, for example, prefer to use the NIOSH Lifting Equation or psychophysical tables. Another practical constraint includes the frequency of analysis performed by the ergonomist or ergonomists within an organisation because the initial costs of acquiring hardware or software may be prohibitive relative to the frequency of their future use. With these factors in mind, we provide an overview of the choices available.

### SURVEY OF CERTIFIED PROFESSIONAL ERGONOMISTS

Dempsey et al. (2005) performed a survey of CPEs certified by the Board of Certification in Professional Ergonomics. Most ergonomists were from the United States, which is a limitation of the generalisability of the results, but the trends for basic tools should be generalisable. The survey was initiated to help structure the contents of a model ‘tool kit’ taking into account the preferences of practising ergonomists and to provide an understanding of the situations they typically face. The intention was that the contents would include typical measurement devices, software, etc. which would accommodate a high percentage of ergonomists.

Table 16.1 provides an overview of the most commonly reported basic tools, observational techniques and direct measurement techniques by the sample of 308 CPEs that responded to the postal survey. The basic tools category had a number of tools that were reported by a majority of respondents, whereas the observational techniques and direct measurements showed less widespread congruence. The observational techniques have a strong representation by techniques to assess MMH tasks. Body discomfort maps were used by over half the sample.

**TABLE 16.1**

**Most Commonly Reported Basic Tools, Observational Techniques and Direct Measurement Techniques and Percentage of Respondents Who Reported Using Such Tools**

| Tool/Technique Classification |     |                                          |     |                         |     |
|-------------------------------|-----|------------------------------------------|-----|-------------------------|-----|
| Basic Tools                   |     | Observational Techniques                 |     | Direct Measurements     |     |
| Tape measure                  | 96% | NIOSH Lifting Equation <sup>a</sup>      | 83% | Grip dynamometer        | 67% |
| Video camera                  | 96% | Biomechanical model                      | 73% | Push/pull force sensor  | 61% |
| Stopwatch                     | 88% | Psychophysical MMH data                  | 73% | Pinch dynamometer       | 53% |
| Digital camera                | 87% | Body discomfort map                      | 56% | Heart rate monitor      | 41% |
| Laptop computer               | 82% | Rapid Upper Limb Assessment <sup>b</sup> | 52% | Instrumented hand tools | 32% |
| Spring gauge                  | 72% | Energy expenditure model <sup>c</sup>    | 44% | Electromyography        | 31% |

Source: Adapted from Dempsey, P.G. et al., *Appl. Ergon.*, 36, 489, 2005.

<sup>a</sup> Waters et al. (1993).

<sup>b</sup> McAtamney and Corlett (1993).

<sup>c</sup> Garg et al. (1978).

## **JOB DEMAND ASSESSMENT TOOLKIT**

Based on the survey discussed in the previous section, a tape measure, stopwatch, device to take photos and videos, a spring gauge to measure weights and forces and a computer (portable computing formats lighter and easier to use in the field than laptops are much more common since the survey was taken) should be included in a toolkit. The results of the observational techniques survey show that, broadly speaking, methods to assess MMH tasks and upper-extremity-intensive tasks are required as well as a body discomfort map. More specific guidance on different choices for assessment tools is provided later in this chapter. With the exception of a biomechanical model, the observational techniques are all available in the literature, and many can be located by web searches. The direct measurements all require devices that must be purchased.

## **INTERVIEWS AND QUESTIONNAIRES**

The first level of any E/HF analysis should include informal interviews with the personnel performing the job of interest as well as management. Operators should explain how they perform tasks, which may or may not coincide exactly with how standard operating procedures, production engineering analyses or management have specified performance. Operators are generally very forthcoming about many aspects of their job, especially with respect to what may be causing discomfort or injuries. Anecdotally, the authors have encountered cases where operators were reluctant to talk about their job or volunteer to have their job performance observed or analysed because of tension with management. Conversely, the opposite has been observed in facilities where safety and health were recognised as organisational values. This crucial aspect of an analysis can certainly be described as more art than science, as experience is very helpful when performing these initial steps of an analysis (consistent with the notion of ‘craft’ discussed in Chapter 1 of this book).

The discomfort surveys discussed here are often useful during these initial informal interviews. Whether the discomfort survey indicates the perceived source of discomfort or not, the ergonomist will begin to understand the problem and can ask operators to discuss the source(s) of discomfort. Many operators are proud of their work, and having someone attempt to understand their job and potential E/HF shortcomings is often viewed positively.

## **DIRECT OBSERVATION**

Direct observation techniques (see Chapter 4) comprise the majority of analyses carried out by many ergonomists. These techniques include any that the ergonomist carries out by observing or measuring different aspects of the work system either in real time or by viewing video or photos captured at the workplace. This information, and especially videos and photos, can be especially helpful for reports or presentations of the findings to those responsible with decision making on the recommendations.

### **Temporal Characteristics**

Time is a critical aspect of work organisation when assessing physical job demands. For repetitive work, cycle time is often used, sometimes in conjunction with duty cycle. Duty cycle is the amount of time within a cycle that the operator is exerting force. Examples include the amount of time grasping products while packing boxes or the amount of time a knife is in contact with meat while performing a specific meat-cutting operation. For MMH tasks, the analyst often needs to know the frequency of a particular handling task per minute.

The choice of measurement instrument should be driven by the precision requirements of the intended analysis and the practical issues of collecting the information from the workplace. For example, when collecting frequency of an MMH task or repetitive assembly task, it is usually possible to use a watch or stopwatch and count the number of repetitions per minute. When performing

a more detailed analysis of the elements of a repetitive assembly task, using video to be able to slow motion is required. The proliferation of video and related software in different platforms has expanded the possibilities dramatically compared to the days of recording video on cassettes and related media.

## Video

In addition to the use of video for capturing the temporal aspects of work, video is extremely useful for estimating postures, performing detailed breakdowns of jobs and tasks into elements, and recording other aspects of the job that require the ergonomist's attention (e.g. visual and auditory interfaces, safety hazards). Although video is extremely useful, taking poor quality video or not capturing all the needed information can limit the extent of analysis, especially when the site is not easily accessed. There are several potential issues with video recording that can be mitigated through practice and proper planning including access and video quality concerns. Access may be an issue when recording in areas where space is limited. Mounting video cameras on the floor or suspending them overhead may help to improve access but will require prior planning to ensure cameras are aimed appropriately. It may also be necessary to use several cameras to allow differing views in tighter spaces. Poor quality video may be the result of poor video capture methods. Shaky video or frequent pans can be disconcerting to the analyst and may be easily resolved with practise. Purchasing video cameras with built-in image stabilisation may also reduce shaky video due to using hand-held cameras or mounting cameras on vibrating surfaces. Environmental factors such as poor lighting, changes in lighting conditions and high levels of dust may also reduce the quality of video available. The inexperienced ergonomist should consider practising their video skills and determining whether the work environment is conducive to video analysis due to any environmental limitations.

Aside from these practical issues, there are a number of important technical considerations. Regardless of the assessment approach, the representativeness of a given analysis is rarely discussed. Efforts to observe multiple operators should be made. When assessing posture and especially when estimating joint angles, the ideal situation is to record video in the plane perpendicular to the plane in which motion occurs (e.g. place video capture plane parallel to sagittal plane when observing trunk flexion). If possible, use multiple videos to reduce parallax errors when multiple angles are observed.

Paquet et al. (2006) recommend determining whether using video-based job analysis will provide the required information before carrying out the analysis. Due to potential limitations in the reliability of quantitative estimates made from video, prioritising job changes made from video may not be advisable. The following recommendations are provided as a basic set of guidelines to maximise the value of video captured in the workplace and the validity of the associated analysis:

1. Use a tripod (or at least a monopod) to minimise motion.
2. Take video in multiple planes to reduce parallax errors and permit more accurate assessment of posture. Multiple angles do not need to be taken simultaneously if more than one video device is not available.
3. Exposures in self-paced manufacturing may vary with time and across operators which will require multiple observations (Paquet et al., 2006).
4. Do not use video analysis to identify peak exposures unless they are expected in advance to occur frequently (Paquet et al., 2006).
5. For repetitive manufacturing where exposure variability is a concern, a strategy of using at least three videos from multiple workers on different days is a more effective strategy than recording multiple videos of a single operator within 1 day (Paquet et al., 2006).
6. Pilot test the video strategy to determine whether it will provide reliable exposure information or if the exposure is highly variable (Paquet et al., 2006).

## DIRECT MEASUREMENT

A main advantage of direct measurement is that continuous and accurate data can often be readily recorded. Because the readings are measured directly while the operator performs their job, variability within and across analysts is not an issue. Direct measurement devices include those devices which can be placed on the person to measure motion, limb accelerations, muscle activity, force or pressure application or other person-specific factors. Additionally, direct measurement devices may also be placed on work surfaces to measure force or pressure applications, frequency of task performance or other factors that may be independent of the worker. Due to the technology used in some direct measurement devices such as accelerometers or gyroscopes, the accuracy and reliability of the device may be deemed unacceptable when used over an extended period of time due to drift, creep or stress relaxation which will be described later in this chapter. However, many products on the market are sufficient for measuring typical work tasks. Issues of exposure variability across time and operators and related issues are the same as observational methods, although direct measurement eliminates potential variability due to intra- and inter-rater variability.

Direct measurement tends to be more expensive and may require more expertise than observational techniques. For example, electromyography (EMG) is complex and requires basic knowledge of anatomy, muscle electrophysiology as well as signal processing techniques. Electrogoniometers, on the other hand, can be purchased with all the hardware and software required to obtain joint angle data with a reasonable amount of system familiarisation. Due to their cost, direct measurement devices are typically purchased with the intention of frequent use. Several examples will be discussed in the following sections.

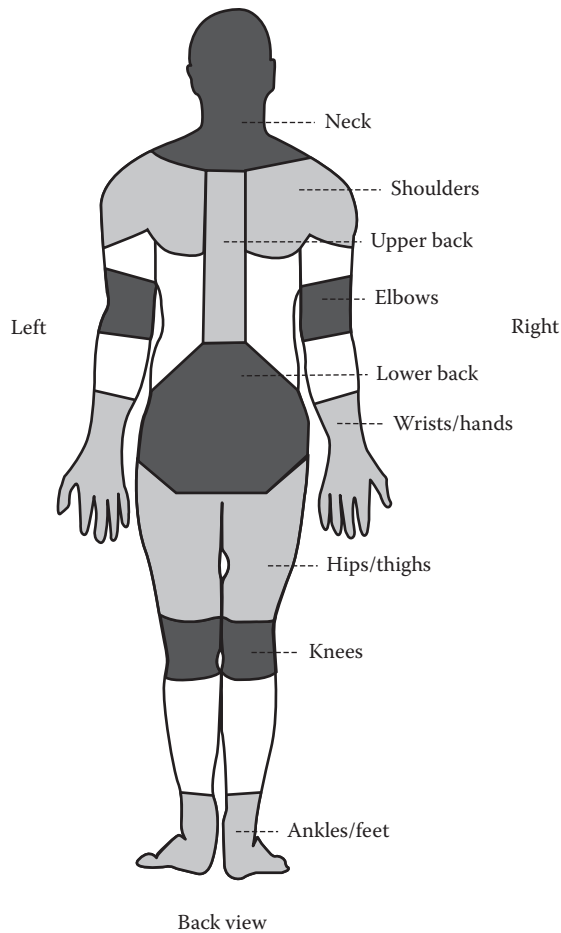
## ASSESSMENT TOOLS AND METHODS

This section provides an overview of the tools and methods used to assess physical demands in the workplace. As discussed in the previous section, factors such as the nature of the problem under investigation, equipment cost, expertise and preference of the ergonomist, and practical constraints such as access to the workplace influence the particular analysis approach. Our goal is to provide an overview of available choices that covers a range of expertise from novice to experienced ergonomist. Because our goal is evaluating work for the prioritisation of redesign and quantification of risks, research methods requiring laboratory simulations or invasive measurements are not covered in this section.

## DISCOMFORT

Discomfort is often considered a precursor to more serious pain or MSDs (Hamberg-van Reenen et al., 2008; Kourinka and Forcier, 1995; Punnett and Wegman, 2004). Mapping areas of discomfort allows ergonomists to determine which parts of the body may be strained during work tasks allowing the ergonomist to focus on those tasks and activities that are likely causing immediate hazards. Questionnaires and body part maps are commonly employed to ascertain areas of discomfort. The investigation of ice cream shops by Dempsey et al. (2000) is an example of how discomfort reports have been used as part of a broader investigation. The discomfort data were considered together with the observational and direct measurement approaches. The discomfort surveys corroborated the findings that show that the frequent and forceful ice cream scooping is the likely cause of the high prevalence of distal upper-extremity discomfort.

A discomfort survey used extensively is that developed by Corlett and Bishop (1976). The authors demonstrated a linear relationship between the growth in discomfort (reported on a five- or seven-point scale) and the time spent performing a static hold until an unbearable level of discomfort was reached. Using this principle, static work can be evaluated by mapping discomfort at several intervals throughout the workday. Workers are asked to identify the areas of discomfort



**FIGURE 16.1** Body map used for evaluating discomfort.

from a body map (Figure 16.1) and provide a ranking on the intensity of the discomfort on a five or seven-point scale. Generally, the scale is anchored at 0 being ‘no discomfort’ and the fifth or seventh point being ‘extreme discomfort’. Average discomfort scores are then plotted against time to show increased discomfort and the extent of recovery (if any) occurring during the day due to rest periods or breaks in exposures. This discomfort mapping can also be used independent of the rating scales when time is limited. In these cases, workers may be asked to rank discomfort areas by first identifying the most uncomfortable areas, followed by the next most uncomfortable areas. Identified areas are given numerical discomfort ratings with 0 being no discomfort, the last reported site having a 1 discomfort rating and the second to last site having a 2 discomfort rating, etc. These ratings are also averaged and plotted against time, allowing identification of the most heavily loaded parts of the body.

The Nordic Musculoskeletal Questionnaire (originally used by Kuorinka et al., 1987 and later modified by Dickinson et al. 1992) is used to provide full body discomfort reporting. It includes the upper and lower back, neck, and upper and lower limbs and determines if symptoms were present during the past 12 months and the last 7 days. Table 16.2 shows the Nordic questionnaire’s basic format for an analysis of the whole body. Separate Nordic questionnaires may also be used for targeting specific areas of the body. An example of a targeted Nordic questionnaire for the low back is shown in Table 16.3. The Nordic questionnaire has the flexibility to be used during a structured interview or to be filled out directly by an employee. However, increased reporting of discomfort

**TABLE 16.2**  
**Nordic Questionnaire**

| Trouble with Locomotive Organs                                                         |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
|----------------------------------------------------------------------------------------|----------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|----------------------------------------------------------|---------------------------------|
| Have you at any time during the last 12 months had trouble (ache, pain, discomfort) in |                                                          | To be answered only by those who have had trouble                                                                                             |                                 |                                                          |                                 |
|                                                                                        |                                                          | Have you at any time during the last 12 months been prevented from doing your normal work (at home or away from home) because of the trouble? |                                 | Have you had trouble at any time during the last 7 days? |                                 |
| Neck                                                                                   |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes                          | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |
| Shoulders                                                                              |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes, in the right shoulder   | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |
|                                                                                        | 3. <input type="checkbox"/> Yes, in the left shoulder    |                                                                                                                                               |                                 |                                                          |                                 |
|                                                                                        | 4. <input type="checkbox"/> Yes, in both shoulders       |                                                                                                                                               |                                 |                                                          |                                 |
| Elbows                                                                                 |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes, in the right elbow      | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |
|                                                                                        | 3. <input type="checkbox"/> Yes, in the left elbow       |                                                                                                                                               |                                 |                                                          |                                 |
|                                                                                        | 4. <input type="checkbox"/> Yes, in both elbows          |                                                                                                                                               |                                 |                                                          |                                 |
| Wrists/hands                                                                           |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes, in the right wrist/hand | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |
|                                                                                        | 3. <input type="checkbox"/> Yes, in the left wrist/hand  |                                                                                                                                               |                                 |                                                          |                                 |
|                                                                                        | 4. <input type="checkbox"/> Yes, in both wrists/hands    |                                                                                                                                               |                                 |                                                          |                                 |
| Upper back                                                                             |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes                          | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |
| Low back (small of back)                                                               |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes                          | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |
| One or both hips/thighs                                                                |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes                          | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |
| One or both knees                                                                      |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes                          | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |
| One or both ankles/feet                                                                |                                                          |                                                                                                                                               |                                 |                                                          |                                 |
| 1. <input type="checkbox"/> No                                                         | 2. <input type="checkbox"/> Yes                          | 1. <input type="checkbox"/> No                                                                                                                | 2. <input type="checkbox"/> Yes | 1. <input type="checkbox"/> No                           | 2. <input type="checkbox"/> Yes |

Source: From Kuorinka, I. et al., *Appl. Ergon.*, 18(3), 233, 1987.

has been shown when this questionnaire is administered as part of targeted efforts to reduce MSD risk factors instead of part of a general health and wellness survey (Andersson et al., 1987). There are numerous published studies that have used this questionnaire, so the ergonomist can compare results as a benchmark. Diverse examples include surveys of cleaners (Chang et al., 2012), a large ( $n = 3003$ ) population sample (Widanarko et al., 2012) and a sample of employees who remove oysters from their shells (Hsu et al., 2011).

## POSTURE AND MOTION

Posture can be static or dynamic, and analysis methods will vary depending on the nature of the work tasks and associated postural durations. In addition to estimating postures from photos and video, several other tools can be used to measure joint postures.



**TABLE 16.3****Nordic Low Back Questionnaire**

1. Have you ever had low back trouble (ache, pain or discomfort)?

1. ☐ No                      2. ☐ Yes

If you answered no to Question 1, do not answer Questions 2–8.

2. Have you ever been hospitalised because of low back trouble?

1. ☐ No                      2. ☐ Yes

3. Have you ever had to change jobs or duties because of low back trouble?

1. ☐ No                      2. ☐ Yes

4. What is the total length of time that you have had low back trouble during the last 12 months?

1. 0 days  
2. 1–7 days  
3. 8–30 days  
4. More than 30 days

If you answered 0 days to Question 4, do not answer Questions 5–8.

5. Has low back trouble caused you to reduce your activity during the last 12 months?

a. Work activity (at home or away from home)?

1. ☐ No                      2. ☐ Yes

b. Leisure activity?

1. ☐ No                      2. ☐ Yes

6. What is the total length of time that low back trouble has prevented you from doing your normal work (at home or away from home) during the last 12 months?

1. 0 days  
2. 1–7 days  
3. 8–30 days  
4. More than 30 days

7. Have you been seen by a doctor, physiotherapist, chiropractor or other such person because of low back trouble during the last 12 months?

1. ☐ No                      2. ☐ Yes

8. Have you had low back trouble at any time during the last 7 days?

1. ☐ No                      2. ☐ Yes

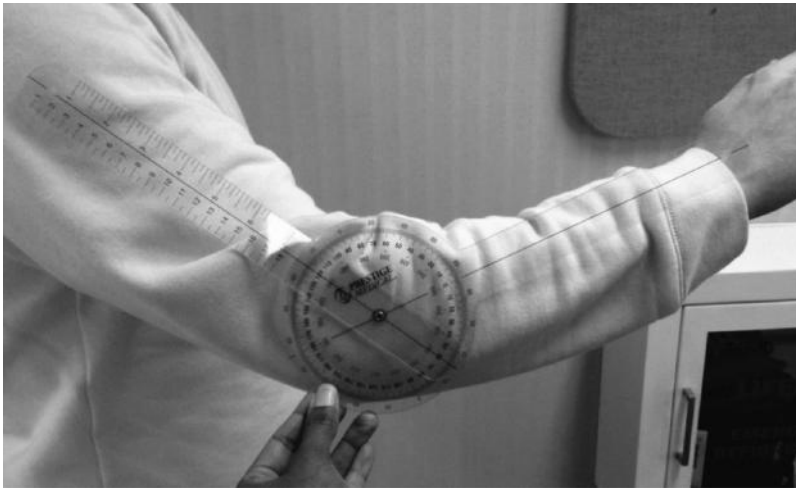
*Source:* From Kuorinka, I. et al., *Appl. Ergon.*, 18(3), 233, 1987.

**Goniometer**

Goniometers are used to measure joint angles and are available in several varieties. The most basic goniometer is comparable to a swing arm protractor and uses two arms with a centre wheel to measure the angle between the two arms (Figure 16.2). These goniometers are commonly used in clinical settings but may also be used when studying static postures to determine included angles of joints such as the elbows and knees.

Electrogoniometers are electrical devices commonly used to provide real-time dynamic angle measures in multiple planes and can be either twin or single axis. Single-axis electrogoniometers are designed to measure angles in a single plane, and twin-axis electrogoniometers are capable of simultaneously measuring angles in two planes. Whether single or twin axis, all electrogoniometers use an electromechanical system to determine joint angles.

Potentiometric electrogoniometers have potentiometers at the centre of joint rotation. As the joint angle changes, the electrical resistance varies, allowing an estimation of joint angles. Typically, these are quite bulky and have limited practical usage due to the fixed position of the joint centre of rotation and their ability to measure only in one plane. Flexible electrogoniometers use strain gauge springs connected to end blocks (Figure 16.3). The end blocks are attached to the body with



**FIGURE 16.2** Basic goniometer.



**FIGURE 16.3** Flexible electrogoniometer.

double-sided tape. As the end blocks move, the resistance of the strain gauge changes, allowing an estimate of joint angle. Similarly, torsionmeters are flexible electrogoniometers that are used to measure the angle of torsion at a joint. They are commonly worn at the neck and back. Flexible electrogoniometers are the most popular goniometers and provide more flexibility over potentiometric goniometers. They are lightweight, very portable, virtually unrestrictive when worn and are available in a variety of sizes for varying joints. However, they require a wired connection to a data logger or a data read-out system, thus limiting their usage in some environments. In rugged work environments such as those where an employee will be likely to get wet or dirty, the flexible electrogoniometers may not be well suited.

### **Industrial Lumbar Motion Monitor**

The Industrial Lumbar Motion Monitor (iLMM) is an externally worn, exoskeleton device designed by researchers from the Biodynamics Laboratory at the Ohio State University's Institute for Ergonomics; this device was developed to assess the dynamic component of occupationally related low-back disorders (Marras et al., 1992). The device is essentially a triaxial electrogoniometer that is affixed at the shoulders and pelvis to trace and measure the motions of the spine. It measures position, velocity and acceleration of the spine in the sagittal, lateral and twisting planes. This device uses a series of potentiometers to measure axial twisting, flexion/extension and lateral rotation of the back. Since its initial construction, the iLMM has undergone several redesigns and is currently on its third version (iLMM3). This new version is lighter and less bulky, allows for wireless tracking and has quicker, simpler setup than previous versions.

## Optoelectronic Systems

Optoelectronic systems (also known as motion capture systems) are used in many settings to determine joint angles. These are video systems that utilise cameras to track markers placed on the body. These markers can be a variety of types including infrared markers, light-emitting diodes or reflective shapes (which are commonly used in passive motion capture systems). The three-dimensional positions of these markers are tracked, and with the use of provided software (and sometimes custom-written analysis scripts), joint angles can be determined. Although extremely precise, these systems can be expensive and require time-consuming calibration and analysis routines. Optical motion capture systems also require a large unobstructed view which is not usually feasible in all work environments. Chapter 17 (by Dempsey and Munro) discusses simulation methods that have been used to generate kinematics when such sophisticated systems are not practical.

## Accelerometry

Accelerometers sense acceleration and its direction. In general, they measure acceleration in multiples of gravity (g). They can also be used to measure angles through a linearisation technique. When on a level surface and oriented such that the z-axis is upwards, an accelerometer would report 1 g of acceleration. When oriented such that the z-axis is downwards, it would report -1 g. Therefore, using linear analysis methods, the multiple-axis joint angles can be estimated from the magnitude of accelerations reported by the accelerometer when mounted correctly on the body. This method is relatively new and requires accurate placement on the body to ensure proper interpretation of results. One benefit is that accelerometers can be wireless and thus allow a completely untethered means of collecting postural data. Data can then be collected and viewed in real time or stored on a data logger. Similarly, gyroscopes measure the rate of rotation and also require numerical integration to produce the desired information. A major limitation with the application of accelerometers or gyroscopes is that they are prone to noise and drift. This drift occurs when small errors are integrated with respect to time, resulting in an amplification of the error.

## Inertial Measurement Units

Inertial measurement units (IMUs) are electronic devices composed of accelerometers and gyroscopes used to determine the acceleration, rotation and velocity of an object. IMUs are commonly used in navigation and tracking to supplement global positioning system (GPS) data and determine object orientation. More recently, IMUs have been used in the study of posture and motion in sports-related research, gait, orthopaedics and biomechanics. IMUs have been utilised to provide accurate measurements of 3D knee angles (Favre et al., 2008), determination of gait temporal events and spatial parameters (Hanlon and Anderson, 2009; Kose et al., 2011; Li et al., 2010), trunk orientation during various tasks (Mazzà et al., 2012; Wong and Wong, 2006) and a variety of other applications. When utilising IMUs in a workplace setting, proper placement and calibration of the sensors are essential. IMUs must be placed such that their outputs are in clinically interpretable angles or analysis should be performed to output results in meaningful coordinate frames. A functional calibration technique has been proposed for 3D knee angle calculation, which may be applied to the angle or elbow (Favre et al., 2004). These systems will likely increase in popularity over time as costs decrease. Today, many smart phones and other portable electronic devices now include IMUs to improve gameplay and GPS systems to track position.

When used in combination with GPS data, video and ground reaction forces, IMUs can become an essential element of a novel motion capture system – called a fusion motion capture system. In this system, data from multiple sources are fused and integrated to quantify segmental as well as whole-body kinetic and kinematics in applications where traditional optical motion capture may be insufficient (Brodie et al., 2008). This allows application of motion capture techniques in environments outside of research laboratories and provides more real-world exposure assessments. As with accelerometers and gyroscopes, a major limitation with the application of an IMU is the drift

associated with accumulated error. This drift can be reduced when using IMUs along with other components such as GPS (when available) or drift reduction algorithms (Borenstein et al., 2009; Cavallo et al, 2005; Ojeda and Borenstein, 2007).

## MANUAL MATERIALS HANDLING

MMH tasks are typically assessed using biomechanical, psychophysical or physiological criteria. The basic approach to selecting a criterion is that biomechanics is most appropriate for low-frequency tasks requiring exertion of high forces, physiology is most appropriate for high-frequency tasks involving light loads or low forces, and psychophysics is most appropriate for intermediate frequencies. Psychophysical tables typically cover a broad range of frequencies, and one approach is to modify the tables to accommodate the biomechanical and physiological criteria (Mital et al., 1997). The NIOSH Lifting Equation (Waters et al., 1993) was designed to provide lifting limits that comply with all three types of criteria. The choice of a particular assessment method can be driven by several influences including any particular regulatory constraints as well as the preferences and expertise of the analyst. Some of these constraints will be highlighted within the specific context of the tool.

### Revised NIOSH Lifting Equation (United States)

At least in the United States, the NIOSH Lifting Equation is one of the most widely used observational assessment tools (see Table 16.1). The equation has drawn interest from other countries as well such as in the Netherlands (Health Council of the Netherlands, 1995) and the United Kingdom (Pinder and Frost, 2011). A thorough description of the equation, its underlying assumptions, restrictions on application and examples are beyond the scope of this chapter, but Waters et al. (1994) provide a comprehensive treatment. However, the basic history and use of the equation are described next.

NIOSH published the original equation for setting weight limits for lifting tasks in 1981 (NIOSH, 1981). The 1981 equation applies to symmetric lifting tasks and was later revised to accommodate asymmetric lifting tasks as well as tasks with less than optimal hand-to-object couplings (Waters et al., 1993, 1994). In addition to the coupling and asymmetry changes, modifications to the equation included a 17 kg reduction of the load constant, modifications to the horizontal multiplier, modifications to the effect of frequency and replacement of multiple limits (the action limit and the maximum permissible limit) by a single limit (recommended weight limit (RWL)). This revised NIOSH Lifting Equation is used to determine the RWL and the Lifting Index (LI) for a specified lifting task.

The RWL is the product of the load constant (23 kg), the multipliers for horizontal location (H) and vertical location (V), distance lifted or lowered (D), angle of asymmetry (A) calculated as indicated by entering each parameter in the respective multiplier, and the frequency and coupling multipliers which are table lookups using frequency (Table 16.4) and coupling characteristics (Tables 16.5 and 16.6 and Figure 16.4):

$$RWL = 23 \text{ kg} \times \left[ \frac{25}{H} \right] \times [1 - 0.003|V - 75|] \times \left[ 0.82 + \frac{4.5}{D} \right] \times FM \times [1 - 0.0032 \times A] \times CM$$

The RWL is calculated at the origin of the lift/lower for lifts/lowers not requiring significant control and at both the origin and destination when significant control is required. Waters et al. (1994) define significant control as requiring precision placement of the load such that the worker needs to regrasp the load near the destination, momentarily hold the load near the destination, or position or guide the load at the destination. When RWL is calculated at both origin and destination, the lower value is used for the assessment.

**TABLE 16.4**  
**Frequency Multiplier Table**

| Frequency, Lifts per Minute (F) | Work Duration |                |                     |                |                     |                |
|---------------------------------|---------------|----------------|---------------------|----------------|---------------------|----------------|
|                                 | $\leq 1$ h    |                | $>1$ but $\leq 2$ h |                | $>2$ but $\leq 8$ h |                |
|                                 | V < 75 cm     | V $\geq$ 75 cm | V < 75 cm           | V $\geq$ 75 cm | V < 75 cm           | V $\geq$ 75 cm |
| $\leq 0.2$                      | 1.00          | 1.00           | 0.95                | 0.95           | 0.85                | 0.85           |
| 0.5                             | 0.97          | 0.97           | 0.92                | 0.92           | 0.81                | 0.81           |
| 1                               | 0.94          | 0.94           | 0.88                | 0.88           | 0.75                | 0.75           |
| 2                               | 0.91          | 0.91           | 0.84                | 0.84           | 0.65                | 0.65           |
| 3                               | 0.88          | 0.88           | 0.79                | 0.79           | 0.55                | 0.55           |
| 4                               | 0.84          | 0.84           | 0.72                | 0.72           | 0.45                | 0.45           |
| 5                               | 0.80          | 0.80           | 0.60                | 0.60           | 0.35                | 0.35           |
| 6                               | 0.75          | 0.75           | 0.50                | 0.50           | 0.27                | 0.27           |
| 7                               | 0.70          | 0.70           | 0.42                | 0.42           | 0.22                | 0.22           |
| 8                               | 0.60          | 0.60           | 0.35                | 0.35           | 0.18                | 0.18           |
| 9                               | 0.52          | 0.52           | 0.30                | 0.30           | 0.00                | 0.15           |
| 10                              | 0.45          | 0.45           | 0.26                | 0.26           | 0.00                | 0.13           |
| 11                              | 0.41          | 0.41           | 0.00                | 0.23           | 0.00                | 0.00           |
| 12                              | 0.37          | 0.37           | 0.00                | 0.21           | 0.00                | 0.00           |
| 13                              | 0.00          | 0.34           | 0.00                | 0.00           | 0.00                | 0.00           |
| 14                              | 0.00          | 0.31           | 0.00                | 0.00           | 0.00                | 0.00           |
| 15                              | 0.00          | 0.28           | 0.00                | 0.00           | 0.00                | 0.00           |
| $>15$                           | 0.00          | 0.00           | 0.00                | 0.00           | 0.00                | 0.00           |

*Source:* Waters, T.R. et al., Applications manual for the revised NIOSH lifting equation, Pub. No. 94-110, DHHS(NIOSH), Cincinnati, OH, 1994.

Once the RWL is calculated, the LI is calculated by dividing the actual load lifted/lowered by the RWL. Waters et al. (1993) suggest that tasks with an LI  $> 1$  pose an increased risk of low back pain for a fraction of the workforce. For jobs with multiple tasks (such as palletising), Waters et al. (1994) describe calculations for the Composite Lifting Index (CLI). This procedure involves similar calculations to the LI calculation but has special provisions for dealing with frequency. For the LI and CLI, spreadsheets or software can be very helpful.

### 1992 Manual Handling Assessment Charts (United Kingdom)

The HSE's regulations publication (HSE, 2004) provides guidance on implementing the 1992 Manual Handling Operations Regulations. Unlike the narrower NIOSH Lifting Equation, the guidance covers a broader set of materials handling situations and also is less specific in suggesting elements such as individual loads. The Manual Handling Assessment Charts, called the MAC (HSE, 2003), were designed to help health, and safety inspectors assess the most common risk factors with the goal of identifying and then reducing the overall risk of the task. The charts also include specific instructions for carrying out MAC assessment for lifting/lowering, carrying and team handling tasks.

A MAC assessment requires that the analyst observe the task to ensure that the task being evaluated is a representative task, select the appropriate (or multiple) types of assessment and follow the steps as outlined in the assessment documentation. Scoring of an assessment is achieved through numerical scores and four colour bands. Green indicates a low level of risk where special risk groups should be considered when appropriate. Amber indicates a medium level of risk where tasks should be examined closely. Red indicates a high level of risk where

**TABLE 16.5**  
**Hand-to-Container Coupling Classification Table**

| Good                                                                                                                                                                                                                                                                  | Fair                                                                                                                                                                                                                          | Poor                                                                                                                                        |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| For containers of optimal design, such as boxes or crates, a 'good' hand-to-object coupling would be defined as handles or hand-hold cut-outs of optimal design [see notes 1–3].                                                                                      | For containers of optimal design, a 'fair' hand-to-object coupling would be defined as handles or hand-hold cut-outs of less than optimal design [see notes 1–5].                                                             | Containers of less than optimal design or loose parts or irregular objects that are bulky, hard to handle or have sharp edges [see note 5]. |
| For loose parts or irregular objects, which are not usually containerised, such as castings, stock and supply materials, a 'good' hand-to-object coupling would be defined as a coupling grip in which the hand can be easily wrapped around the object [see note 6]. | For containers of optimal design with no handles or hand-hold cut-outs or for loose parts or irregular objects, a 'fair' hand-to-object coupling is defined as a grip in which the hand can be flexed about 90° [see note 4]. | Lifting non-rigid bags (i.e. bags that sag in the middle).                                                                                  |

*Source:* Waters, T.R. et al., Applications manual for the revised NIOSH lifting equation, Pub. No. 94-110, DHHS(NIOSH), Cincinnati, OH, 1994.

*Notes:*

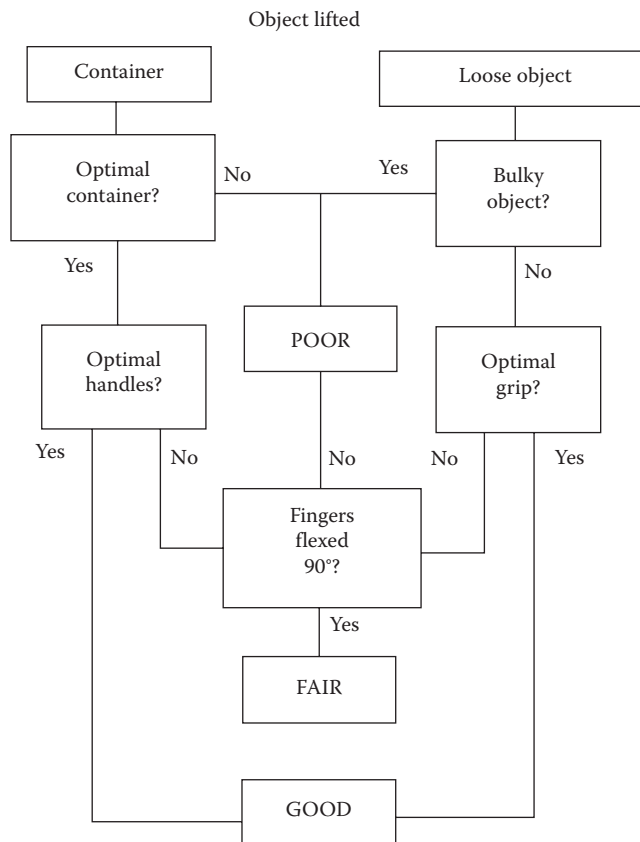
1. An optimal handle design has 0.75–1.5 in. (1.9–3.8 cm) diameter, ~4.5 in. (11.5 cm) length, 2 in. (5 cm) clearance, cylindrical shape and a smooth non-slip surface.
2. An optimal hand-hold cut-out has the following approximate characteristics: ~1.5 in. (3.8 cm) height, 4.5 in. (11.5 cm) length, semi-oval shape, ~2 in. (5 cm) clearance, smooth nonslip surface and ~0.25 in. (0.60 cm) container thickness (e.g. double thickness cardboard).
3. An optimal container design has ~16 in. (40 cm) frontal length, ~12 in. (30 cm) height and a smooth non-slip surface.
4. A worker should be capable of clamping the fingers at nearly 90 in. under the container, such as required when lifting a cardboard box from the floor.
5. A container is considered less than optimal if it has a frontal length >16 in. (40 cm), height >12 in. (30 cm), rough or slippery surfaces, sharp edges, asymmetric centre of mass, unstable contents or requires the use of gloves. A loose object is considered bulky if the load cannot easily be balanced between the hand-grasps.
6. A worker should be able to comfortably wrap the hand around the object without causing excessive wrist deviations or awkward postures, and the grip should not require excessive force.

**TABLE 16.6**  
**Coupling Multiplier Table**

| Coupling Type | Coupling Multiplier |           |
|---------------|---------------------|-----------|
|               | V < 75 cm           | V ≥ 75 cm |
| Good          | 1.00                | 1.00      |
| Fair          | 0.95                | 1.00      |
| Poor          | 0.90                | 0.90      |

*Source:* Waters, T.R. et al., Applications manual for the revised NIOSH lifting equation, Pub. No. 94-110, DHHS(NIOSH), Cincinnati, OH, 1994.





**FIGURE 16.4** Decision tree for coupling quality. (From Waters, T.R. et al., Applications manual for the revised NIOSH lifting equation, Pub. No. 94-110, DHHS(NIOSH), Cincinnati, OH, 1994.)

prompt action is needed. Finally, purple indicates a very high level of risk where close scrutiny must be applied to the task to reduce injury risks.

Similar risk factors are used for the differing tasks as shown in Tables 16.7 through 16.10 and Figures 16.5 and 16.6. Once a numerical score and colour band are determined for each risk factor, they are recorded on the MAC score sheet as shown in Figure 16.7. All risk factor scores are added to determine a total score. The total score is then used to prioritise tasks for further review, and the colour bands are used to determine which areas of the tasks should receive the most attention.

### Psychophysical Tables

The psychophysical approach to designing MMH tasks is based on tables of loads and forces that accommodate different percentages of the male and female populations. The values are provided for various frequencies, distances for carries, pushes and pulls, and similar task parameters. The most widely recognised and most thorough tables are those reported by Snook and Ciriello (1991). These tables were developed (using psychophysics techniques – see Chapter 4) over several decades and represent data from a significant sample of industrial workers. They cover lifting, lowering, pushing, pulling and carrying. The goal of the analyst should be to accommodate a high percentage of the population. The tables are used by collecting task parameters (e.g. hand distance from body,

**TABLE 16.7****Colour Bands and Numerical Scores for Lifting/Lowering Operations Using the MAC**

| Risk Factor                                  | Green Colour Band                                                                                         |   | Amber Colour Band                                                                                          |   | Red Colour Band                                                             |   |
|----------------------------------------------|-----------------------------------------------------------------------------------------------------------|---|------------------------------------------------------------------------------------------------------------|---|-----------------------------------------------------------------------------|---|
| A: Load weight/<br>frequency                 | See Figure 16.5 for colour bands and scores                                                               |   |                                                                                                            |   |                                                                             |   |
| B: Hand distance<br>from the lower<br>back   | CLOSE: Upper arms<br>aligned vertically and<br>upright trunk                                              | 0 | MODERATE: Upper arms<br>angled away from body<br>or trunk bent forward                                     | 3 | FAR: Upper arms angled<br>away from body and<br>trunk bent forward          | 6 |
| C: Vertical lift<br>region                   | Above knee and/or below<br>elbow height                                                                   | 0 | Below knee and/or above<br>elbow height                                                                    | 1 | Floor level or below or<br>at head or above                                 | 3 |
| D: Trunk twisting<br>and sideways<br>bending | No bending or twisting                                                                                    | 0 | Trunk twists in relation to<br>the hips and thighs or<br>worker leans to one side<br>as the load is lifted | 1 | Trunk both twists and<br>bends to the side as the<br>load is lifted         | 2 |
| E: Postural<br>constraints                   | Movement is unhindered                                                                                    | 0 | Adopts restricted posture<br>due to space or<br>workstation design                                         | 1 | Posture is severely<br>restricted                                           | 3 |
| F: Grip on load                              | Containers with<br>well-designed handles or<br>handholds or loose parts<br>enabling a<br>comfortable grip | 0 | Container with poor<br>handles or handholds or<br>fingers clamped at 90°<br>under container                | 1 | Containers of poor<br>design or non-rigid<br>sacks or unpredictable<br>load | 2 |
| G: Floor surfaces                            | Dry and clean floor in<br>good condition                                                                  | 0 | Dry floor but in poor<br>condition, worn or<br>uneven                                                      | 1 | Contaminated/wet or<br>steep, sloping floor or<br>unstable footing          | 2 |
| H: Other<br>environmental<br>factors         | No other factors (e.g.<br>extreme temperature,<br>strong air movement,<br>extreme lighting)               | 0 | One environmental factor<br>present                                                                        | 1 | Two or more factors<br>present                                              | 2 |

*Source:* Adapted from HSE, *Manual Handling: Manual Handling Operations Regulations 1992 (as Amended), Guidance on Regulations L23*, 3rd edn., HSE, Sudbury, Ontario, Canada, 2004.

frequency and lifting range for lifting and lowering) and then using the tables to find the percentage of the population accommodated. The tables provide weights or forces that accommodate different percentages of the population. The analyst often interpolates using the measured weight or force to find the percentage of the population accommodated.

Overall, the psychophysical approach is very useful, and there is fairly extensive data on MMH tasks available. For a more detailed overview of the approach, including a detailed discussion of advantages and limitations as well as available data sources, the reader is referred to Dempsey (2006) and Dempsey and Maynard (2005). Mital et al. (1997) provide psychophysical tables modified to also comply with biomechanical and physiological criteria because participants sometimes select loads that conflict with spinal compression and energy expenditure criteria. Mital et al. also cover materials handling in unusual postures and one-handed handling tasks.

### Physiological Assessment

Measuring energy expenditure is expensive and can be difficult in the workplace. The equipment required is typically not available to the ergonomist, but an alternative approach is to use regression models that predict energy expenditure from task, operator and workplace parameters.

**TABLE 16.8****Colour Bands and Numerical Scores for Carrying Operations Using the MAC**

| Risk Factor                          | Green Colour Band                                                                             |   | Amber Colour Band                                                                  |   | Red Colour Band                                                     |   |
|--------------------------------------|-----------------------------------------------------------------------------------------------|---|------------------------------------------------------------------------------------|---|---------------------------------------------------------------------|---|
| A: Load weight/frequency             | See Figure 16.6 for colour bands and scores                                                   |   |                                                                                    |   |                                                                     |   |
| B: Hand distance from the lower back | CLOSE: Upper arms aligned vertically and upright trunk                                        | 0 | MODERATE: Upper arms angled away from body or trunk bent forward                   | 3 | FAR: Upper arms angled away from body and trunk bent forward        | 6 |
| C: Asymmetrical trunk/load           | Load and hands symmetrical in front of the trunk                                              | 0 | Load and hands asymmetrical, upright body position                                 | 1 | One handed carrying to the individual's side                        | 2 |
| D: Postural constraints              | Movement is unhindered                                                                        | 0 | Adopts restricted posture during the carry                                         | 1 | Posture is severely restricted                                      | 3 |
| E: Grip on load                      | Containers with well-designed handles or handholds or loose parts enabling a comfortable grip | 0 | Container with poor handles or handholds or fingers clamped at 90° under container | 1 | Containers of poor design or non-rigid sacks or unpredictable loads | 2 |
| F: Floor surfaces                    | Dry and clean floor in good condition                                                         | 0 | Dry floor but in poor condition, worn or uneven                                    | 1 | Contaminated/wet or steep sloping floor or unstable footing         | 2 |
| G: Other environmental factors       | No other factors (e.g. extreme temperature, strong air movement, extreme lighting)            | 0 | One environmental factor present                                                   | 1 | Two or more factors present                                         | 2 |
| H: Carry distance                    | 2–4 m                                                                                         | 0 | 4–10 m                                                                             | 1 | ≥ 10 m                                                              | 3 |
| I: Obstacles en route                | No obstacles or carry route is flat                                                           | 0 | Steep slope or trip hazards or up steps                                            | 2 | Carrying up ladders                                                 | 3 |

*Source:* Adapted from HSE, *Manual Handling: Manual Handling Operations Regulations 1992 (as Amended), Guidance on Regulations L23*, 3rd edn., HSE, Sudbury, Ontario, Canada, 2004.

As Table 16.1 shows, the models developed by Garg et al. (1978) are still used quite frequently, at least in the United States. The equations are too detailed to discuss in detail here, but the basic premise of the Garg et al. (1978) equations is that energy expenditure is assumed to be additive, and energy expenditure for each component of a job can be estimated and the total demand determined. An advantage of these equations is that they cover a wide range of materials handling activities and situations. Although the assumption of additivity adds to the robustness, the assumption has been found to lead to large errors (see Dempsey et al., 2008 for a more detailed discussion).

Dempsey et al. (2008) developed a set of regression models to predict energy expenditure during the performance of individual (lift, lower, push, pull, carry) and combination (lift-carry-lower, palletising, depalletising) materials handling tasks from a representative sample of experienced male and female materials handlers. The chosen combination tasks were commonly encountered (Dempsey, 2003), and developing models specific to the performance of the tasks as a unit rather than estimating from adding the costs of the individual lift, carry and lower components alleviates the additivity problem discussed earlier. A fairly large sample (19 males and

**TABLE 16.9****Colour Bands and Numerical Scores for Team Handling Operations Using the MAC**

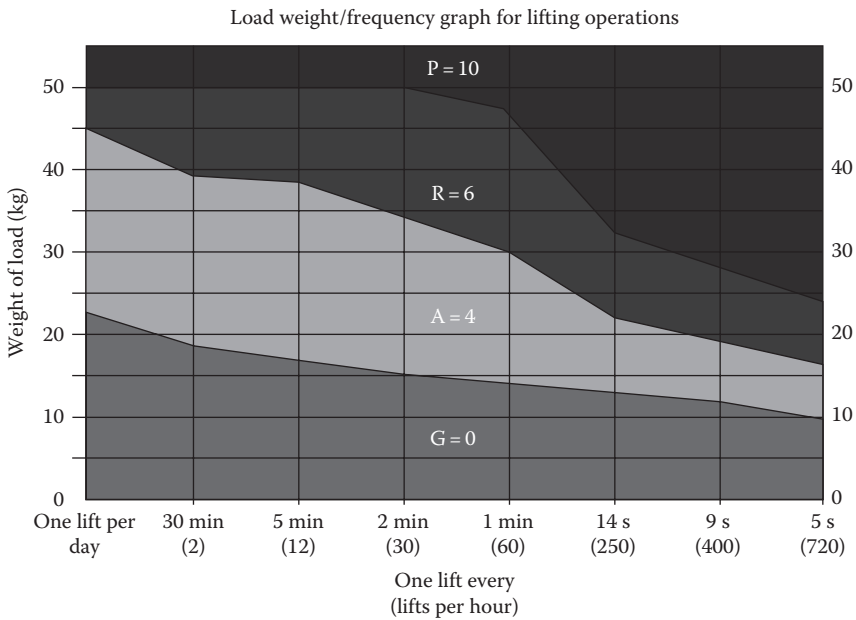
| <b>Risk Factor</b>                   | <b>Green Colour Band</b>                                                                      |   | <b>Amber Colour Band</b>                                                           |   | <b>Red Colour Band</b>                                              |   |
|--------------------------------------|-----------------------------------------------------------------------------------------------|---|------------------------------------------------------------------------------------|---|---------------------------------------------------------------------|---|
| A: Load weight                       |                                                                                               |   | See Table 16.10                                                                    |   |                                                                     |   |
| B: Hand distance from the lower back | CLOSE: Upper arms aligned vertically and upright trunk                                        | 0 | MODERATE: Upper arms angled away from body or trunk bent forward                   | 3 | FAR: Upper arms angled away from body and trunk bent forward        | 6 |
| C: Asymmetrical trunk/load           | Load and hands symmetrical in front of the trunk                                              | 0 | Loads and hands asymmetrical, upright body position                                | 1 | One handed carrying to the individual's side                        | 2 |
| D: Postural constraints              | Movement is unhindered                                                                        | 0 | Adopts restricted posture during the carry                                         | 1 | Posture is severely restricted                                      | 3 |
| E: Grip on load                      | Containers with well-designed handles or handholds or loose parts enabling a comfortable grip | 0 | Container with poor handles or handholds or fingers clamped at 90° under container | 1 | Containers of poor design or non-rigid sacks or unpredictable loads | 2 |
| F: Floor surfaces                    | Dry and clean floor in good condition                                                         | 0 | Dry floor but in poor condition, worn or uneven                                    | 1 | Contaminated/wet or steep, sloping floor or unstable footing        | 2 |
| G: Other environmental factors       | No other factors: (extreme temperature, strong air movement, extreme lighting)                | 0 | One environmental factor present                                                   | 1 | Two or more factors present                                         | 2 |
| H: Carry distance                    | 2–4 m                                                                                         | 0 | 4–10 m                                                                             | 1 | ≥10 m                                                               | 3 |
| I: Obstacles en route                | No obstacles or carry route is flat                                                           | 0 | Steep slope or trip hazards or up steps                                            | 2 | Carrying up ladders                                                 | 3 |

Source: Adapted from HSE, *Manual Handling: Manual Handling Operations Regulations 1992 (as Amended), Guidance on Regulations L23*, 3rd edn., HSE, Sudbury, Ontario, Canada, 2004.

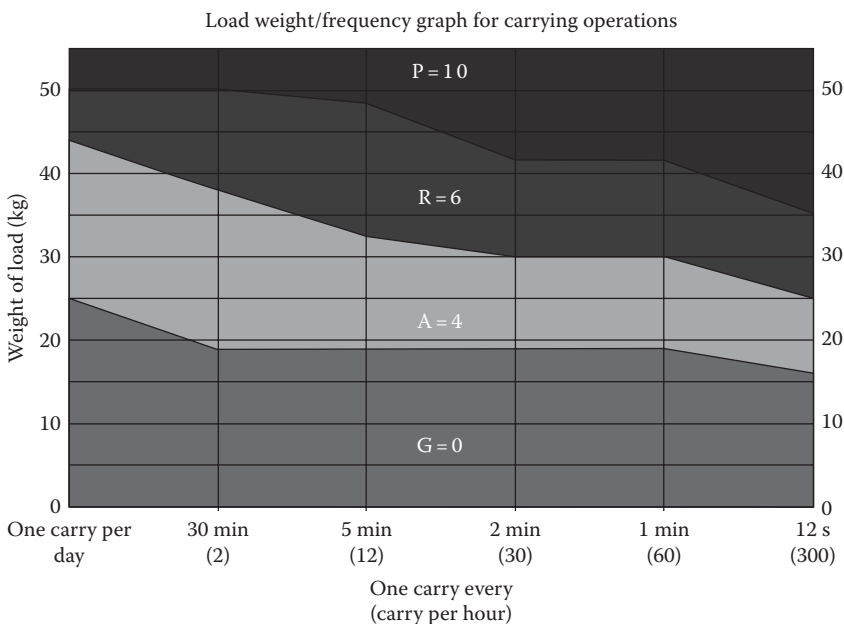
**TABLE 16.10****Colour Bands and Numerical Scores for the Load Weight Risk Factor for Team Handling Operations Using the MAC**

| <b>Green Colour Band</b> |   | <b>Amber Colour Band</b> |   | <b>Red Colour Band</b> |   | <b>Purple Colour Band</b> |    |
|--------------------------|---|--------------------------|---|------------------------|---|---------------------------|----|
| 2 person <35 kg          | 0 | 2 person 30–50 kg        | 4 | 2 person 50–85 kg      | 6 | 2 person >85 kg           | 10 |
| 3 person <40 kg          |   | 3 person 40–75 kg        |   | 3 person 75–125 kg     |   | 3 person >125 kg          |    |
|                          |   | 4 person 40–100 kg       |   | 4 person 100–170 kg    |   | 4 person >170 kg          |    |

Source: Adapted from HSE, *Manual Handling: Manual Handling Operations Regulations 1992 (as Amended), Guidance on Regulations L23*, 3rd edn., HSE, Sudbury, Ontario, Canada, 2004.



**FIGURE 16.5** Load weight/frequency graph for carrying operations using the MAC. For the colour bands G, green; A, amber; R, red and P, purple. (From HSE, *Manual Handling: Manual Handling Operations Regulations 1992 (as Amended), Guidance on Regulations L23*, 3rd edn., HSE, Sudbury, Ontario, Canada, 2004.)



**FIGURE 16.6** Load weight/frequency graph for carrying operations using the MAC. For the colour bands G, green; A, amber; R, red and P, purple. (From HSE, *Manual Handling: Manual Handling Operations Regulations 1992 (as Amended), Guidance on Regulations L23*, 3rd edn., HSE, Sudbury, Ontario, Canada, 2004.)

19 females) with varying anthropometry was used to generate over 8000 values which were used to develop the regression models. These models can be used to compare the energy demands of different tasks.

### Biomechanical Assessment

Like physiological assessment, performing a biomechanical assessment at the workplace can be challenging due to the equipment that is sometimes required. Chapter 17 discusses the motion capture requirements for dynamic analyses which is why most ergonomists either choose to use a static model or a method that estimates biomechanical loading. A model used fairly extensively over the years is the University of Michigan 3D Static Strength Prediction Program™, the basis of which is described by Chaffin et al. (1999) and which has been refined from an earlier two-dimensional model. The model estimates spinal compression as well as the percentage of the population capable of exerting the joint torques required by an activity. This or another commercially available model is often used by practising ergonomists because the level of effort and expertise required to build such a model can be extensive.

An alternative to using a model is predictive equations analogous to those discussed earlier for predicting energy consumption. The ergonomist takes measurements of task parameters and enters them into a regression model to predict spinal compression. Spinal compression is the criterion used to assess alternate designs or to assess whether a task exceeds a criterion such as the 3400 N limit selected by Waters et al. (1993). Potvin (1997) reported regression models to predict spinal compression from NIOSH equation inputs (horizontal and vertical location parameters discussed earlier). Two models developed had high R-squared values (0.989, 0.983), and there was, in general, good agreement between the estimates and the results from a biomechanical model. More recently, van Dieën et al. (2010) compared results from Potvin's (1997) model to data collected in the laboratory using 3D kinematics from a motion capture system and data from force platforms. The Potvin model estimates were approximately 20% lower than estimates from laboratory equipment in conjunction with a biomechanical model. Because rank ordering of tasks was consistent for both methods, van Dieën et al. (2010) concluded that the regression models were sufficient for comparative analyses of different task designs.

### UPPER LIMB ASSESSMENTS

Upper limb disorders (ULDs) are prevalent in many industries where workers perform repetitive tasks (Bernard, 1997; Kourinka and Forcier, 1995; Nordstrom et al., 1997; Rempel and Punnett, 1997; Werner et al., 1997). The RULA and the Assessment of Repetitive Tasks of the upper limbs (the ART Tool) are two commonly used assessment methods. Both methods require observations and communications with workers and provide a relative index to classify the hazards which can be used to aide redesign and prioritise job modifications.

#### RULA

The RULA was developed to assess exposures and determine the action level of tasks thought to contribute to the development of ULDs (McAtamney and Corlett, 1993). This technique considers postures, forces and muscles activity associated with work tasks by examining the upper and lower body postures. The upper body group (Table 16.11) consists of the upper arms, lower arms and wrists; the lower body group (Table 16.12) includes the legs, trunk and neck. A score sheet is used to calculate a grand score for each task based on the sum of the posture, muscle use and force scores (Figure 16.8). Posture is observed and given a numerical classification based on values provided in charts (Figure 16.9). These values are then used to determine a posture score for Tables 16.11 and 16.12. No force or muscle activity measurements are taken; instead, charts are used to determine a score based on the duration of static posture and weight



MAC: Score sheet

Company name: \_\_\_\_\_

Task description: \_\_\_\_\_

Are the indications that the task is high risk?  
(Tick the appropriate boxes)

☐ Task has a history of manual handling incidents

☐ Task is known to be hard work or high risk

☐ Employees doing the work show signs that they are finding it hard work (e.g. breathing heavily, sweating)

☐ Other indications, if so what? \_\_\_\_\_

Signature \_\_\_\_\_ Date \_\_\_\_\_

| Risk factors                                                             | Color band<br>(G, A, R or P) |       |      | Numerical source |       |      |
|--------------------------------------------------------------------------|------------------------------|-------|------|------------------|-------|------|
|                                                                          | Lift                         | Carry | Team | Lift             | Carry | Team |
| Load weight and lift/carry frequency                                     |                              |       |      |                  |       |      |
| Hand distance from the lower back                                        |                              |       |      |                  |       |      |
| Vertical lift region                                                     |                              |       |      |                  |       |      |
| Trunk twisting/sideways bending<br>Asymmetrical trunk/ load carrying     |                              |       |      |                  |       |      |
| Postural constraints                                                     |                              |       |      |                  |       |      |
| Grip on the load                                                         |                              |       |      |                  |       |      |
| Floor surfaces                                                           |                              |       |      |                  |       |      |
| Other environmental factors                                              |                              |       |      |                  |       |      |
| Carry distance                                                           |                              |       |      |                  |       |      |
| Obstacles en route (carrying only)                                       |                              |       |      |                  |       |      |
| Communication and coordination<br>(team handling only)                   |                              |       |      |                  |       |      |
| Other risk factors, e.g. Individual factors, psychological factors, etc. | Total score:                 |       |      |                  |       |      |

**FIGURE 16.7** Score sheet for use with the MAC. (Adapted from HSE, *Manual Handling: Manual Handling Operations Regulations 1992 (as Amended), Guidance on Regulations L23*, 3rd edn., HSE, Sudbury, Ontario, Canada, 2004.)

of loads (Figure 16.10). A table is used to determine the grand score from the subtotal scores for Tables 16.11 and 16.12 (Figure 16.11). This grand score is then used to determine the action level of the associated task.

In the RULA technique, individual scores are assigned due to risks associated with posture, force and muscle activity and thus allow the identification of areas which are driving risks due to less desirable conditions. When using this method for prioritisation of redesign, modifications should be focused on the areas associated with the highest scores. This technique is relatively easy to conduct and provides reliable results without specialised tools or equipment beyond the associated tables and charts.

### ART Tool (United Kingdom)

The ART Tool is designed to assess some common risk factors in repetitive work that contribute to the development of ULDs (HSE, 2010). The ART Tool is well suited for tasks that require the use of the upper limbs, occur at least 1–2 h per day and are performed at least once every few minutes. The ART Tool is comprised of the assessment guide, flow chart, task description form and score sheet. The assessment guide includes an overview of the tool, details on the risk factors it assesses and some criteria for its utilisation. The flow chart provides a clear process for performing a task assessment and can be used as a guide through the analysis process. The task description form (Figure 16.12) is a record of the analysis and used to document the task and any associated exposures. The score sheet (Figure 16.13) is used to record the score and other indicators of risks for the analysis. During the assessment, all risks are scored numerically and by colour. Numerical values vary by the risks, and a consistent colour scheme is used whereby green indicates a low level of risk (and lowest numerical value), amber indicates a medium level of risk where the task should be examined closely and red indicates a high level of risk where prompt action is needed. Intermediate numerical values can be used to classify exposures. The assessment consists of four stages: A, B, C and D. A detailed overview of this process is provided later in this section. Throughout this

**TABLE 16.11**  
**Table Used to Determine Posture Scores for Upper Body Using the RULA Method**

| Upper Limb Posture Score |           |                     |   |       |   |       |   |       |   |
|--------------------------|-----------|---------------------|---|-------|---|-------|---|-------|---|
|                          |           | Wrist Posture Score |   |       |   |       |   |       |   |
|                          |           | 1                   |   | 2     |   | 3     |   | 4     |   |
|                          |           | Twist               |   | Twist |   | Twist |   | Twist |   |
| Upper Arm                | Lower Arm | 1                   | 2 | 1     | 2 | 1     | 2 | 1     | 2 |
| 1                        | 1         | 1                   | 2 | 2     | 2 | 2     | 3 | 3     | 3 |
|                          | 2         | 2                   | 2 | 2     | 2 | 3     | 3 | 3     | 3 |
|                          | 3         | 2                   | 3 | 3     | 3 | 3     | 3 | 4     | 4 |
| 2                        | 1         | 2                   | 3 | 3     | 3 | 3     | 4 | 4     | 4 |
|                          | 2         | 3                   | 3 | 3     | 3 | 3     | 4 | 4     | 4 |
|                          | 3         | 3                   | 4 | 4     | 4 | 4     | 4 | 5     | 5 |
| 3                        | 1         | 3                   | 3 | 4     | 4 | 4     | 4 | 5     | 5 |
|                          | 2         | 3                   | 4 | 4     | 4 | 4     | 4 | 5     | 5 |
|                          | 3         | 4                   | 4 | 4     | 4 | 4     | 5 | 5     | 5 |
| 4                        | 1         | 4                   | 4 | 4     | 4 | 4     | 5 | 5     | 5 |
|                          | 2         | 4                   | 4 | 4     | 4 | 4     | 5 | 5     | 5 |
|                          | 3         | 4                   | 4 | 4     | 5 | 5     | 5 | 6     | 6 |
| 5                        | 1         | 5                   | 5 | 5     | 5 | 5     | 6 | 6     | 7 |
|                          | 2         | 5                   | 6 | 6     | 6 | 6     | 6 | 7     | 7 |
|                          | 3         | 6                   | 6 | 6     | 7 | 7     | 7 | 7     | 8 |
| 6                        | 1         | 7                   | 7 | 7     | 7 | 8     | 8 | 8     | 9 |
|                          | 2         | 8                   | 8 | 8     | 8 | 9     | 9 | 9     | 9 |
|                          | 3         | 9                   | 9 | 9     | 9 | 9     | 9 | 9     | 9 |

**TABLE 16.12**  
**Table Used to Determine Posture Scores for Lower Body Using the RULA Method**

| Neck, Trunk, Legs Posture Score |  |                     |   |      |   |      |   |      |   |      |   |
|---------------------------------|--|---------------------|---|------|---|------|---|------|---|------|---|
|                                 |  | Trunk Posture Score |   |      |   |      |   |      |   |      |   |
|                                 |  | 1                   |   | 2    |   | 3    |   | 4    |   | 5    |   |
|                                 |  | Legs                |   | Legs |   | Legs |   | Legs |   | Legs |   |
| Neck Posture Score              |  | 1                   | 2 | 1    | 2 | 1    | 2 | 1    | 2 | 1    | 2 |
| 1                               |  | 1                   | 3 | 2    | 3 | 3    | 4 | 5    | 5 | 6    | 6 |
| 2                               |  | 2                   | 3 | 2    | 3 | 4    | 5 | 5    | 5 | 6    | 7 |
| 3                               |  | 3                   | 3 | 3    | 4 | 4    | 5 | 5    | 6 | 6    | 7 |
| 4                               |  | 5                   | 5 | 5    | 6 | 6    | 7 | 7    | 7 | 7    | 8 |
| 5                               |  | 7                   | 7 | 7    | 7 | 7    | 8 | 8    | 8 | 8    | 8 |
| 6                               |  | 8                   | 8 | 8    | 8 | 8    | 9 | 8    | 9 | 9    | 9 |

RULA score sheet 1

|                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|
| <p>(a)</p> <div style="border: 1px solid black; padding: 2px;">             Upper arm<br/>             Lower arm<br/>             Wrist<br/>             Wrist twist           </div> | <div style="display: flex; align-items: center; justify-content: space-between;"> <div>             Use table A<br/> <div style="border: 1px solid black; width: 60px; height: 30px; margin: 5px;"></div>             Posture score A           </div> <div>+</div> <div> <div style="border: 1px solid black; padding: 2px;">Muscle use</div> <div style="border: 1px solid black; width: 60px; height: 30px; margin: 5px;"></div> </div> <div>+</div> <div> <div style="border: 1px solid black; padding: 2px;">Force</div> <div style="border: 1px solid black; width: 60px; height: 30px; margin: 5px;"></div> </div> <div>=</div> <div> <div style="border: 1px solid black; width: 60px; height: 30px; margin: 5px;"></div>             Score C           </div> </div> <div style="text-align: center; margin: 10px 0;"> <div style="border: 1px solid black; padding: 2px; display: inline-block;">Left</div> <div style="border: 1px solid black; padding: 2px; display: inline-block;">Right</div> </div> <div style="display: flex; align-items: center; justify-content: center;"> <div>→</div> <div style="border: 1px solid black; padding: 5px; margin: 0 10px;">Use table C</div> <div>→</div> </div> <div style="display: flex; align-items: center; justify-content: center;"> <div>→</div> <div style="border: 1px solid black; padding: 5px; margin: 0 10px;">Grand score</div> </div> | <p>Comments<br/>Task:</p> |
| <p>(b)</p> <div style="border: 1px solid black; padding: 2px;">             Neck<br/>             Trunk<br/>             Legs           </div>                                        | <div style="display: flex; align-items: center; justify-content: space-between;"> <div>             Use table B<br/> <div style="border: 1px solid black; width: 60px; height: 30px; margin: 5px;"></div>             Posture score B           </div> <div>+</div> <div> <div style="border: 1px solid black; padding: 2px;">Muscle use</div> <div style="border: 1px solid black; width: 60px; height: 30px; margin: 5px;"></div> </div> <div>+</div> <div> <div style="border: 1px solid black; padding: 2px;">Force</div> <div style="border: 1px solid black; width: 60px; height: 30px; margin: 5px;"></div> </div> <div>=</div> <div> <div style="border: 1px solid black; width: 60px; height: 30px; margin: 5px;"></div>             Score D           </div> </div> <div style="display: flex; align-items: center; justify-content: center; margin-top: 10px;"> <div>→</div> <div style="border: 1px solid black; padding: 5px; margin: 0 10px;">Grand score</div> </div>                                                                                                                                                                                                                                                                                                                                                                                                                       |                           |

**FIGURE 16.8** Score sheet for the RULA method. (a) the calculations for the upper body group and (b) the calculations for the lower body group.

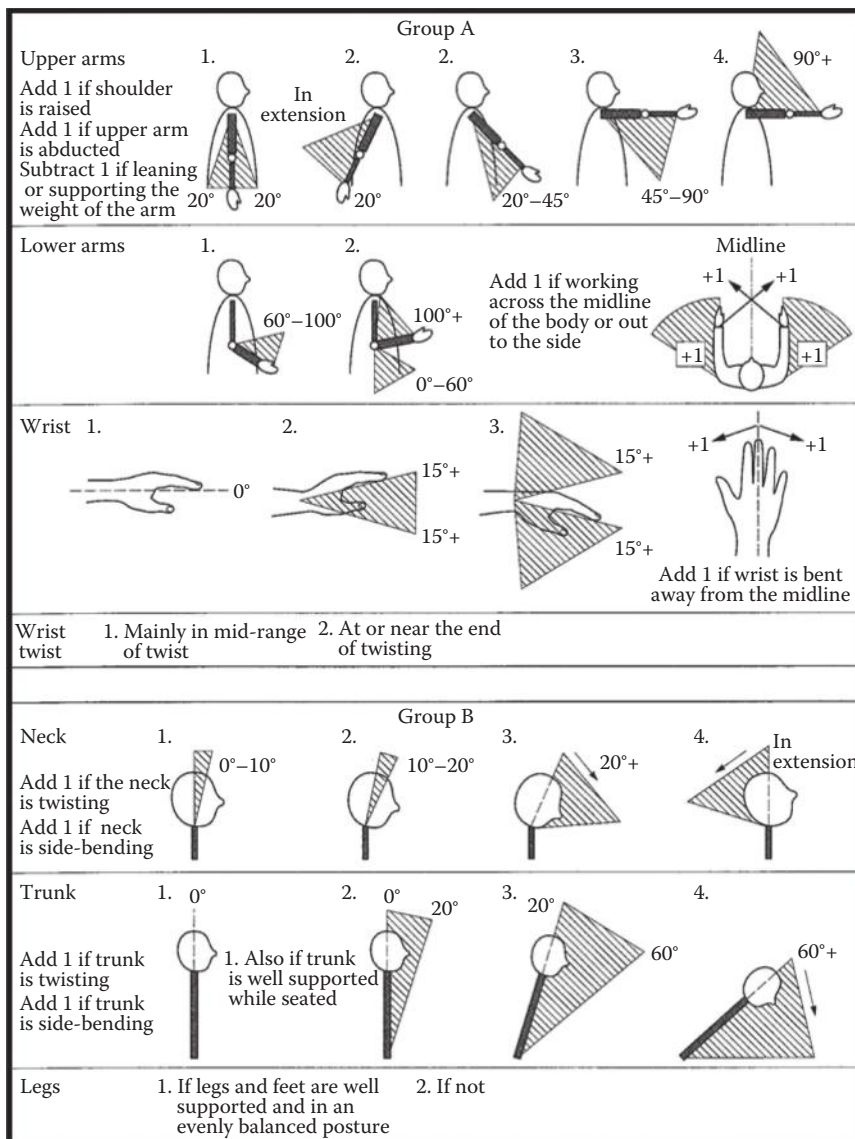
section, hazard colour and numerical scores are presented as a hazard colour (corresponding to a numerical score), for example green (0).

Stage A assesses the frequency and duration of the movements of the arm and the repetition of the arm and hand (both left and right). Arm movements are observed and classified as infrequent, frequent and very frequent. Repetition is determined by observing the motion of the hand and arm (excluding the fingers) over a set time period. Repetition is classified as 10 movements or less per minute, 11–20 times per minute and more than 20 times per minute. The scores assigned are green (0), amber (3) and red (6), respectively.

Stage B examines the force levels required and considers the exerted forces and the duration of exertion. Where possible, workers should report the forces exerted by the hand as light, moderate, strong or very strong. Otherwise, guidance is provided to help classify the force levels (Table 16.13). Numerical values and hazard levels are assigned based on force level and durations using the chart in Figure 16.14. Multiple numerical values are assigned for the amber and red hazard levels.

Stage C considers the postures of the head/neck, back, arm and wrists. For all body parts, neutral postures most of the time (awkward posture <15% of the time) are considered green-level hazards with a numerical value of 0. All other postures are evaluated based on the relative percentage of time spent in awkward postures. Amber-level hazards are assigned when in awkward postures ‘part of the time’, which is defined as between 15% and 30% of the time. Red-level hazards are assigned when awkward postures are observed more than half (>50%) of the time. All amber- and red-level hazards are assigned numerical values of 1 and 2, respectively, except for the arm posture, which has numerical values of 2 and 4. Grip type is observed and used to rate the hazard associated with the hands/finger grip. Power grip has a green hazard level, and pinch or wide finger grips are handled as the other awkward postures. Some guidance on awkward postures is given as provided here (HSE, 2010):

- The neck is considered to be bent or twisted if an obvious angle between the neck and back can be observed as a result of performing the task.
- The back posture is considered awkward if more than 20° of twisting or bending is observed.
- The arm is considered to adopt an awkward posture if the elbow is raised to around chest height and the arm is unsupported (e.g. not resting on a workbench).
- The wrist is considered to be bent or deviated if an obvious wrist angle can be observed.



**FIGURE 16.9** Numerical classifications for postures using the RULA method.

Stage D considers additional factors such as breaks, work pace, other potential hazards and duration of the task. Psychosocial factors which could affect task performance or risks should also be recorded; however, these factors will not contribute to a score. Breaks are taken into account by observing the maximum amount of time a worker performs a task without a significant pause, which is at least 5–10 min. These durations are assigned a hazard level and numerical value. Less than 1 h of continuous work or work with several short breaks is considered green-level hazards with numerical values of 0. Continuous work performed 1–2 h and 2–3 h are considered amber-level hazards with numerical values of 2 and 4, respectively. Continuous work performed 3–4 h and over 4 h are considered red-level hazards with numerical values of 6 and 8, respectively. Employees should be interviewed regarding the pace of their work. If it is not difficult to keep

**Muscle use score**

**Give a score of 1** if the posture is:

- Mainly static, for example, held for longer than 1 min
- Repeated more than four times/min

**Forces or load score**

| 0                                                          | 1                                  | 2                                                     | 3                                                                                                           |
|------------------------------------------------------------|------------------------------------|-------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| No resistance or less than 2 kg intermittent load or force | 2–10 kg intermittent load or force | 2–10 kg static load<br>2–10 kg repeated load or force | 10 kg or more static load<br>10 kg or more repeated loads or forces<br>Shock or forces with a rapid buildup |

**FIGURE 16.10** Tables used to determine muscle and forces scores for the RULA method.

|                         |   | Score D (neck, trunk, legs) |   |   |   |   |   |    |
|-------------------------|---|-----------------------------|---|---|---|---|---|----|
|                         |   | 1                           | 2 | 3 | 4 | 5 | 6 | 7+ |
| Score C<br>(upper limb) | 1 | 1                           | 2 | 3 | 3 | 4 | 5 | 5  |
|                         | 2 | 2                           | 2 | 3 | 4 | 4 | 5 | 5  |
|                         | 3 | 3                           | 3 | 3 | 4 | 4 | 5 | 6  |
|                         | 4 | 3                           | 3 | 3 | 4 | 5 | 6 | 6  |
|                         | 5 | 4                           | 4 | 4 | 5 | 6 | 7 | 7  |
|                         | 6 | 4                           | 4 | 5 | 6 | 6 | 7 | 7  |
|                         | 7 | 5                           | 5 | 6 | 6 | 7 | 7 | 7  |
|                         | 8 | 5                           | 5 | 6 | 7 | 7 | 7 | 7  |

**Table C** Grand score table

- Action level 1** A score of one or two indicated that posture is acceptable if it is not maintained or repeated for long periods
- Action level 2** A score of three or four indicates further investigation is needed and changes may be required
- Action level 3** A score of five or six indicates investigation and changes are required soon
- Action level 4** A score of seven or more indicates investigation and changes are required immediately

**FIGURE 16.11** Grand score table used to determine the action level for the RULA method.

up with the work, this is a green (0)-level hazard. If it is sometimes difficult to keep up with the work, this is an amber (1)-level hazard. If it is often difficult, this is a red (2)-level hazard. Other potential factors are also scored at the discretion of the rater based on the number of factors present. When no other factors are present, the hazard is scored as green (0) level. When one factor is present, it is scored as amber (1) level, and when two or more factors are present, it is scored as red (2) level. The overall duration of the task over the full course of the workday or shift, excluding breaks, is also considered. Work duration is used to determine a duration multiplier which will be multiplied by the task score to yield a final exposure score. These multipliers are 0.5 for tasks less than 2 h, 0.75 for tasks between 2 and 4 h, 1 for tasks between 4 and 8 h and 1.4 for tasks longer than 8 h.

| ART Task description form                                                                                                                           |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|--|--|--|
| Assessor Name:                                                                                                                                      | Date:                                                                                   |  |  |  |  |  |  |  |  |  |  |  |  |
| Company Name:                                                                                                                                       | Location:                                                                               |  |  |  |  |  |  |  |  |  |  |  |  |
| Name of Task:                                                                                                                                       |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| Task description:                                                                                                                                   |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| Weight of any items handled:<br><i>If items weigh more than 8 kg and the tasks involves manual handling consider using the MAC</i>                  |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| Which side of the body is primarily involved?                                                                                                       | Left      Right      Both                                                               |  |  |  |  |  |  |  |  |  |  |  |  |
| What hand tools are used?                                                                                                                           |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| Production rate (If available)                                                                                                                      | Units per shift, hour or minute (circle as appropriate)                                 |  |  |  |  |  |  |  |  |  |  |  |  |
| How often is the task repeated?                                                                                                                     | Every _____ seconds                                                                     |  |  |  |  |  |  |  |  |  |  |  |  |
| Draw the breaks in the shift                                                                                                                        |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| <table border="1"> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> </table> |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
|                                                                                                                                                     |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| First hour —————→                                                                                                                                   |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| How long does a worker perform the task?                                                                                                            | —without a break      hours<br>—in a typical day or shift (excluding breaks)      hours |  |  |  |  |  |  |  |  |  |  |  |  |
| How often does an individual perform the task? (e.g. daily, weekly, etc.)                                                                           |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| How often is the task carried out within the organisation? (e.g. daily, etc.)                                                                       |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| Do workers rotate to other tasks?                                                                                                                   |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |
| If so what tasks?                                                                                                                                   |                                                                                         |  |  |  |  |  |  |  |  |  |  |  |  |

**FIGURE 16.12** Task description form used with the ART Tool. (Adapted from HSE, *Assessment of Repetitive Tasks of the Upper Limbs (The ART Tool)*, HSE, Sudbury, Ontario, Canada, 2010.)

The final exposure score is evaluated such that low exposures (exposure scores between 0 and 11) should be evaluated on individual circumstances, medium scores (exposure scores between 12 and 21) require further investigations and high scores (exposure scores greater than 21) require urgent investigation. Special considerations are given to tasks where job rotation strategies are employed to reduce cumulative exposures. When task rotations are used, each task performed by the worker should be investigated separately to ensure the worker is not rotating to a task with similar high exposures.

Both RULA and the ART Tool are easily learned and give consistent results with reasonable accuracy. Their effectiveness is dependent upon the expertise of the users in understanding the ergonomics of the working situations under study and their ability to recognise awkward postures and other potential hazards. As with other assessment methods, the effectiveness of RULA can be increased when used as part of a larger intervention programme (McAtamney and Corlett, 1993) and with appropriate work analysis methods (Drury, 1987).

## FORCE AND EXERTIONS

To determine levels of force or physical exertion associated with occupational tasks, ergonomists frequently use several specific tools. These tools can be classified as pressure sensors, force sensors or EMG. This section is intended to give an overview of some of the commonly used equipment in these categories.



| Art Tool Score Sheet                                                             |          |       |           |       |
|----------------------------------------------------------------------------------|----------|-------|-----------|-------|
| Enter the color band and numerical score for each risk factor in the table below |          |       |           |       |
| Risk factors                                                                     | Left arm |       | Right arm |       |
|                                                                                  | Color    | Score | Color     | Score |
| A1 Arm movements                                                                 |          |       |           |       |
| A2 Repetition                                                                    |          |       |           |       |
| B Force                                                                          |          |       |           |       |
| C1 Head/neck posture                                                             |          |       |           |       |
| C2 Back posture                                                                  |          |       |           |       |
| C3 Arm posture                                                                   |          |       |           |       |
| C4 Wrist posture                                                                 |          |       |           |       |
| C5 Hand/finger grip                                                              |          |       |           |       |
| D1 Breaks                                                                        |          |       |           |       |
| D2 Work pace                                                                     |          |       |           |       |
| D3 Other factors                                                                 |          |       |           |       |
| Task Score                                                                       |          |       |           |       |
| D4 Duration Multiplier                                                           |          | ×     |           | ×     |
| Exposure Score                                                                   |          |       |           |       |
| D5 Psychosocial factors                                                          |          |       |           |       |

Are there other indications that the task is high risk?

☐ The task or similar tasks have a history of ULDs (e.g. company accident book, medically diagnosed cases of ULDs)

☐ There are signs workers find the task difficult (e.g. wearing arm supports or bandages, reporting discomfort, aches or pains). Ask the workers if they have any of these symptoms.

☐ Other indications? If so, describe: \_\_\_\_\_

**FIGURE 16.13** Score sheet used with the ART Tool. (Adapted from HSE, *Assessment of Repetitive Tasks of the Upper Limbs (The ART Tool)*, HSE, Sudbury, Ontario, Canada, 2010.)

**TABLE 16.13**  
**Descriptions of Force Levels for Use with Stage B of the ART Tool**

|                   |                                                                                                                                                                                                                                                                                                                                |
|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Light force       | There is no indication of any particular effort                                                                                                                                                                                                                                                                                |
| Moderate force    | Force need to be exerted. For example: <div>Pinching or gripping objects with some effort</div> <div>Moving levers or pushing buttons with some effort</div> <div>Manipulating lids or components with some effort</div> <div>Pushing or forcing items together with some effort</div> <div>Using tools with some effort</div> |
| Strong force      | Force is obviously high, strong or heavy                                                                                                                                                                                                                                                                                       |
| Very strong force | Force is near to the maximum level that the worker can apply                                                                                                                                                                                                                                                                   |

Source: From HSE, *Assessment of Repetitive Tasks of the Upper Limbs (The ART Tool)*, HSE, Sudbury, Ontario, Canada, 2010.

|  |                                         | Worker's description of the level of force exerted with the hand |          |                   |                   |
|--|-----------------------------------------|------------------------------------------------------------------|----------|-------------------|-------------------|
|  |                                         | Light                                                            | Moderate | Strong            | Very strong       |
|  | Infrequent                              | G0                                                               | A1       | R6                | Changes required* |
|  | Part of the time<br>(15%–30%)           | G0                                                               | A2       | R9                | Changes required* |
|  | Almost half<br>the time<br>(40%–60%)    | G0                                                               | A4       | R12               | Changes required* |
|  | Almost all<br>the time<br>(80% or more) | G0                                                               | R8       | Changes required* | Changes required* |

\*Changes to the task are required due to unacceptable levels of force

**FIGURE 16.14** Recommendations associated with Stage B of the ART Tool. For the colour bands, G, green; A, amber and R, red. (From HSE, *Assessment of Repetitive Tasks of the Upper Limbs (The ART Tool)*, HSE, Sudbury, Ontario, Canada, 2010.)

### Pressure Sensors

Pressure measurement systems are typically used to identify areas of high-contact pressures when using hand tools or in seating, quantifying gripping pressures and determining foot pressures in gait and fall prevention research.

#### *In-Shoe Systems*

In-shoe pressure measurement systems have been used to quantify pressures at the foot–shoe interface in gait and fall prevention research. Two popular in-shoe pressure measurement insole systems have been compared in the literature: the Pedar® system (Novel, St Paul, MN) and the F-scan® system (Tekscan, Boston, MA). Both systems give pressure maps showing areas of higher and lower pressures. The Pedar system uses capacitive sensors, while the F-scan system uses resistive technology. One study found the Pedar system to have the greatest accuracy and precision when measurements were taken under specific conditions (Hsiao et al., 2002).

#### *Hand-Worn Systems*

Hand-worn systems are commonly used to provide measures of gripping pressure. These systems may utilise a glove with pressure sensors located on the finger tips and palms or may include wearable finger sleeves and palm attachments. The Grip™ System (Tekscan, Boston, MA) uses tactile

pressure sensors which can be built into a glove or worn directly on the hand. Similarly, the Pliance® System (Novel, St. Paul, MN) has glove sensor attachments which can be worn on the hand or built into a glove. This system includes five small strips to measure the pressure distribution at the hand-machine interface. The Finger TPS II, Tactile Pressure Sensor system (Pressure Profile Systems Inc., Los Angeles, CA), uses capacitive sensor technology to quantify the pressures at the fingers and palm through a system with wearable finger sleeves and palm sensor pad.

### *Pressure Mats and Strips*

Pressure mats and strips offer more versatility in usage compared to in-shoe or hand-worn systems. These systems offer the ability to instrument a wide variety of tools or equipment. In addition to the glove sensor, the Pliance System (Novel, St Paul, MN) offers a wide variety of sensors which can be used together or independently. Cylinder sensors are specially designed for measuring pressures when gripping a cylindrical object; strip sensors consist of a single row of sensors for usage in smaller areas; and hand mat sensors may be used to measure pressures in specific hand or finger regions. The I-Scan System (Tekscan, Boston, MA) consists of thin durable sensors which can also be used in a variety of applications.

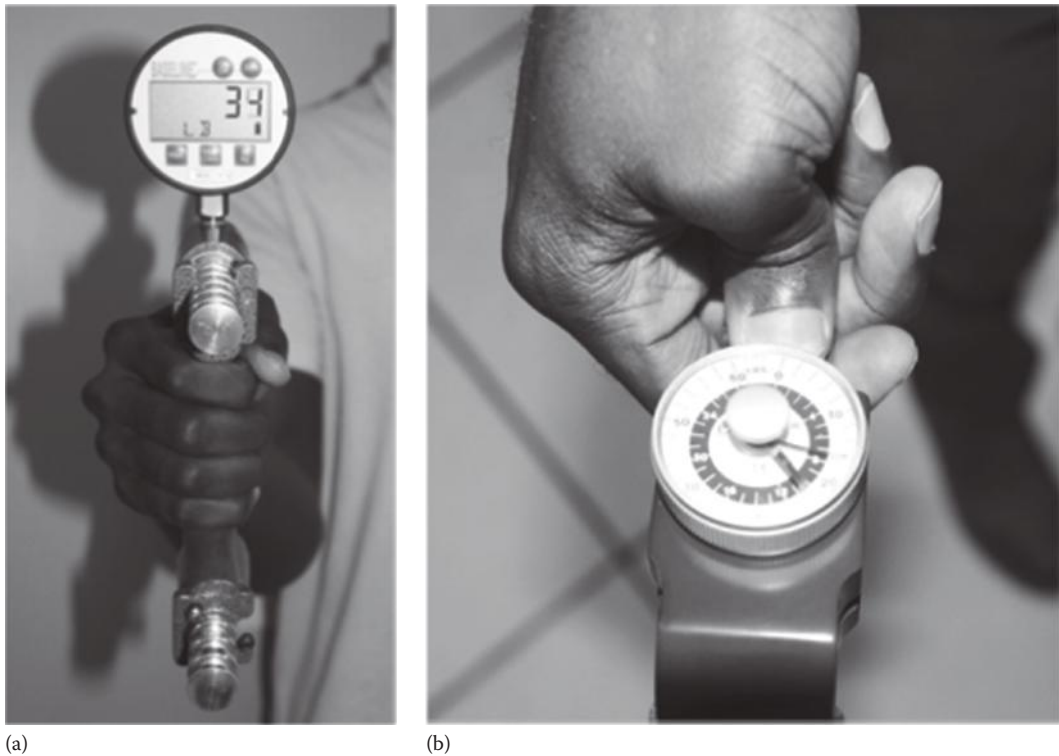
### *Limitations*

Common issues with pressure measurement systems depend on the type of technology used. For workplace exposure assessments, there are three main types of technologies: capacitive, resistive and piezoresistive. All pressure sensors work the same way. There is an electrical element connected to a diaphragm whose electrical properties are affected by the change in pressure applied to the diaphragm. Capacitive pressure sensors use a diaphragm fixed to a deformable capacitor plate. Another rigid capacitor plate is fixed to some substrate creating a reference point. Pressure is applied to the diaphragm resulting in the deformation of the capacitor plate. As the distance between the capacitor plates changes, there is a change in capacitance which is correlated to some level of pressure. Resistive pressure sensors use a diaphragm attached to strain gauges as a means to measure pressure. As pressure is applied to the diaphragm, the diaphragm deflects, elongating or compressing the attached strain gauges. This results in a change in the electrical resistance of the strain gauge which is correlated to some level of pressure. Piezoresistive pressure sensors use specialised strain gauges attached to the diaphragm. Piezoresistive strain gauges utilise monocrystalline semiconductor technology which results in higher changes in resistance than standard strain gauges, increasing their sensitivity to minute pressure changes.

When using pressure sensors, several limitations must be considered. Hysteresis (past activity affecting current measures) is a concern when assessing cyclic tasks. Due to hysteresis, pressure readings during loading and unloading will not show the same pattern of pressure. Stress relaxation and creep are tendencies of viscoelastic materials subjected to constant stress. Over time, there may be a decrease in the pressure readings which are not due to a decrease in pressure application (stress relaxation) or to an increase in pressure readings which are not due to an increased pressure application (creep). Using piezoresistive pressure, sensors may reduce these effects compared to capacitive and resistive sensor technologies. Users of pressure sensors should be cautioned that the factor being measured is pressure and not force. Using pressure sensors to determine applied force is not advised and will likely result in misrepresentations of force levels. In addition, all sensors have a stringent calibration routine which must be followed to ensure accurate measurements.

### **Force Sensors**

Force measurement systems are used in workplace assessments for a variety of reasons. Forces may be the necessary inputs for biomechanical models or may be necessary for other repetitive work assessments. A force sensor is any device which converts a mechanical force into an electrical output signal. Several force sensors are routinely used for workplace assessments including dynamometers, strain gauges and force platforms.



**FIGURE 16.15** Power grip (a) and pinch grip (b) hand dynamometers.

### *Hand Dynamometers*

A hand dynamometer is a device used to measure grip strength. In many cases, this device is used to measure the maximum voluntary contraction (MVC) strength of power or pinch grips (Figure 16.15). Most dynamometers include a gauge directly on the unit to indicate maximal force achieved. Some dynamometers are intended for more continuous study and measure forces in real time, thus requiring a data logger or tethering to a computer.

### *Strain Gauges*

In an ergonomic assessment of jobs requiring repetitive gripping, repetition and duration, data should be supplemented with estimations of applied grip forces. Grip force may be directly measured through instrumented tools or may be estimated through replications of force using a dynamometer. Accuracy of force replication may be compromised and varies highly among people (Casey et al., 2002). When possible, direct measurement of forces via instrumented tools or equipment should be used. A strain gauge is a device used to measure the strain in an object when under deformation. The gauge is attached to the object, and as the gauge is deformed, there is a change in the electrical resistance of the strain gauge which is correlated to force. Metal foil strain gauges are commonly used in workplace assessment as a means to instrument handles of hand tools. To quantify grip force, three strain gauges are mounted to one end of the handle core allowing the measurement of horizontal, vertical and axial moments. For further information on instrumenting hand tools, please see McGorry (2001).

Strain gauges are used in a variety of applications for the measurement of force. When added to the core of hand tools, they provide accurate measures of grip force. When arranged in a wheatstone bridge configuration, they are commonly employed as load cells (a wheatstone bridge is an electrical

circuit configuration whereby two legs of a bridge circuit are balances and used to determine an unknown electrical resistance). Very small changes in resistance are readily detected due to their disruption of the balance in the circuit's resistance. Strain gauges are also the fundamental technology used in force platforms and hand dynamometers. When utilising strain gauges for other applications, it is essential to consider the temperature as most gauges have temperature-dependent changes in electrical resistance which may require compensation during data analysis or may be minimised by purchasing temperature-compensated gauges.

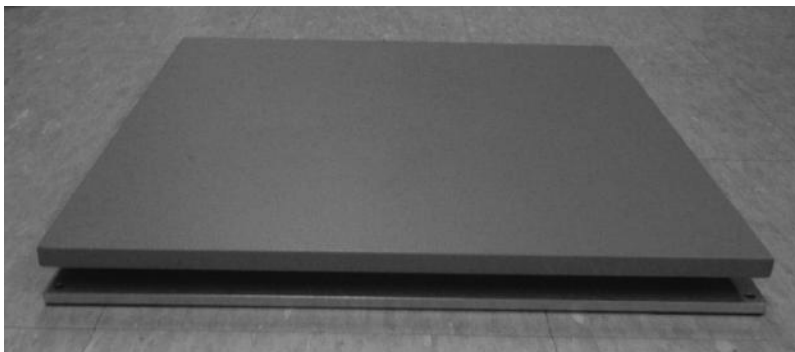
### *Force Platforms*

Force platforms (also known as force plates) are instruments used to measure ground reaction forces. The most commonly used force platforms are triaxial and utilise strain gauge or piezo-electric technology. Forces and moments are measured in three dimensions and are then used to determine the location of the centre of pressure for the applied force. Force platforms are available in permanent mounting as well as portable varieties, with the portable variety generally having a reduced force capacity. A force measurement system typically consists of a force platform, a signal amplifier, a data acquisition card and specialised software provided by the manufacturer. A portable system will have reduced hardware requirements and may require only a force platform and a USB connection into the computer for use with the provided software (Figure 16.16). Multiple analysis software packages are commercially available for the collection and exporting of force data.

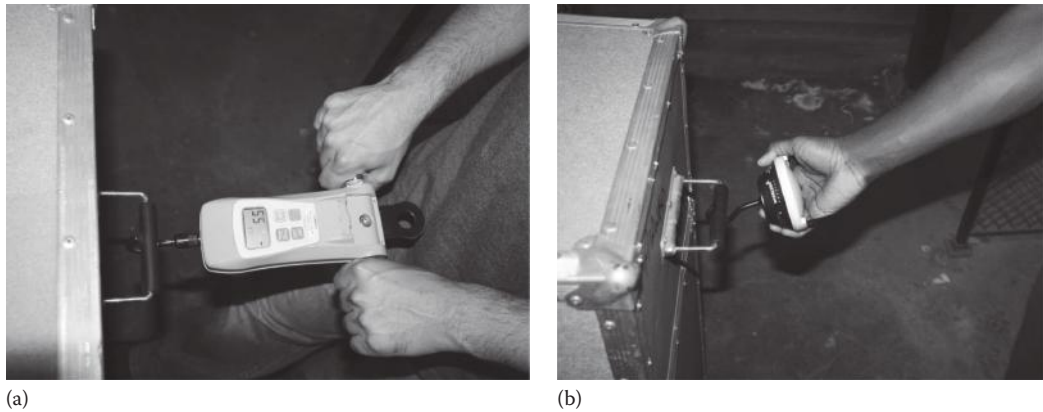
When using force platforms for workplace assessments, proper mounting is essential. Force platforms measure and output data in the coordinate system specified by the platform manufacturer (which can be in a z-down orientation). Any changes in orientation when installed will need to be set up in the platform acquisition software or accounted for in data analysis. In addition, the force platform should be mounted flush with the walking or standing surface. When properly utilised, portable force platforms provide a great means of measuring ground reaction forces for subsequent use in models of workplace exposures.

### *Push/Pull Gauges*

Push/pull gauges are routinely used to quantify push/pull forces in workplace tasks. Some commonly used push/pull gauges include the Chatillon® MSE 100 Ergonomic Gauge (AMETEK, Inc., Largo, FL), the Shimpo Javelin Force Gauge (NIDEC-Shimpo America Corporation, Itasca, IL) and the ergoFET (Hoggan Health Industries, Inc., West Jordan, UT), which comes in a handheld and a two-handled variety (Figure 16.17). Most systems have onboard displays of peak and average force. The ergoFET system uses Bluetooth technology to provide continuous data acquisition.



**FIGURE 16.16** Portable force platform.



**FIGURE 16.17** Commonly used push/pull gauges. Two-handed Shimpo (a) and handheld ergoFET (b).

### Electromyography

EMG determines the level of motor unit activity within a muscle over a period of time by recording the myoelectric signals generated by a muscle when in use. It can be used to compare the relative levels of activity between muscles, determine the onset or presence of muscle fatigue and estimate muscle force.

When assessing the relative activity of a muscle across time or between differing muscles or individuals, it is necessary to normalise the EMG to a reference value. Depending on the purpose of the assessment and the type of data assessed, this reference value can be a full maximum contraction, a submaximal contraction, the mean activity during a trial, the peak activity during the trial or some other reference contraction (Burden and Bartlett, 1999; Perry and Bekey, 1981). Maximum contractions may also be isometric (contractions with a fixed muscle length) or isokinetic/isotonic (contractions with a fixed velocity of motion). In most cases, isometric or isokinetic MVCs are used.

Muscle force and EMG activity are well correlated, and normalised EMG has been used to estimate the amount of force exerted by a muscle. This estimation relies on the relationship between force and muscle activity, which has been documented as a linear (Chaffin et al., 1980; Milner-Brown and Stein, 1975; Perry and Bekey, 1981) as well as a nonlinear relationship depending on the muscle of interest (Woods and Bigland-Ritchie, 1983). Statistical analyses are performed, and regression equations are used to determine the relationship between EMG and force exertion. Force estimations from EMG activity have been performed for the hand muscles (Bao and Silverstein, 2005), knee moments (Lloyd and Besier, 2003) and erector spinae extensor moment (Dolan and Adams, 1993). Whenever possible, forces should be directly measured through force sensors or calculated through a combination of sensors and biomechanical modelling.

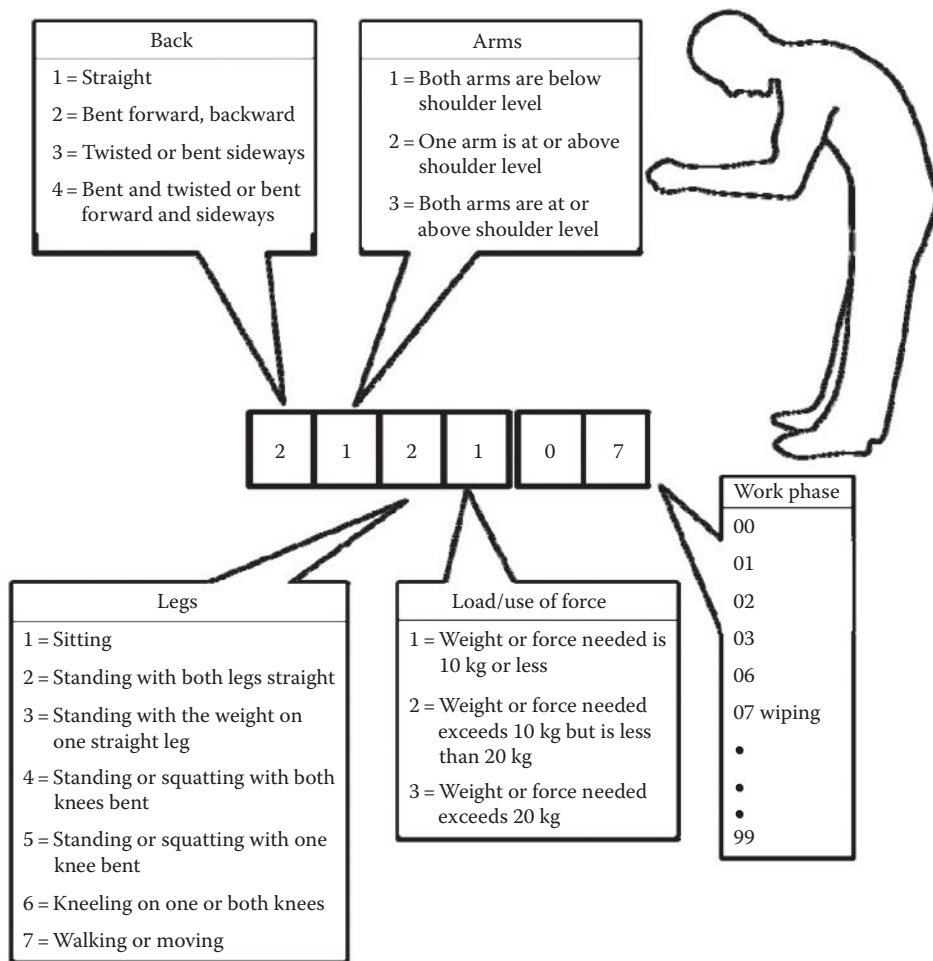
### WHOLE-BODY ASSESSMENT

Whole-body assessment methods are also used to evaluate workplace exposures and prioritise tasks for redesign. The Ovako Working Posture Analysis System (Finnish Institute for Occupational Health, 1992) and the Rapid Entire Body Assessment (REBA) method (Hignett and McAtamney, 2000) are two commonly used tools.

#### Ovako Working Posture Analysis System

One of the first whole-body postural assessment methods was developed by the Ovako Oy Company, in conjunction with the Finnish Institute for Occupational Health (1992), to evaluate the working postures of steelworkers. This system uses direct observations whereby the user observes the work





**FIGURE 16.18** Posture and force codes used in the OWAS method.

being performed, then records the associated back, arm and leg postures, and also the forces and the phases of the work task. Observations are taken every 30–60 s. Using the data from these samples, the work task and its associated work phases can be evaluated. Each posture is given a code along with the magnitude of force used (Figure 16.18). The frequency of the posture is determined, and the proportion of work task spent in this posture is then calculated. Using a table, the action category of the task can be determined from the posture and force codes (Figure 16.19). Considering posture and percentage of working time spent in the associated postures, an action category for the work posture can also be determined from a table (Figure 16.20).

### Rapid Entire Body Assessment

To accommodate the requirement to consider additional factors specific to the health care industry, Hignett and McAtamney (2000) developed REBA. As the name implies, this tool examines the whole body and considers posture, coupling, activity, loading and muscle activity. The body is divided into two major categories: (1) Group B – the arms and (2) Group A – the rest of the body. Group A includes the legs, trunk and neck; and Group B includes the upper arms, lower arms and wrists (Finnish Institute of Occupational Health, 1992). The assessment process involves several steps. First, each segment is given a posture score based on guidelines provided in segment diagrams (Figure 16.21). Next, several tables are used to get to the final score (Figure 16.22). All scores

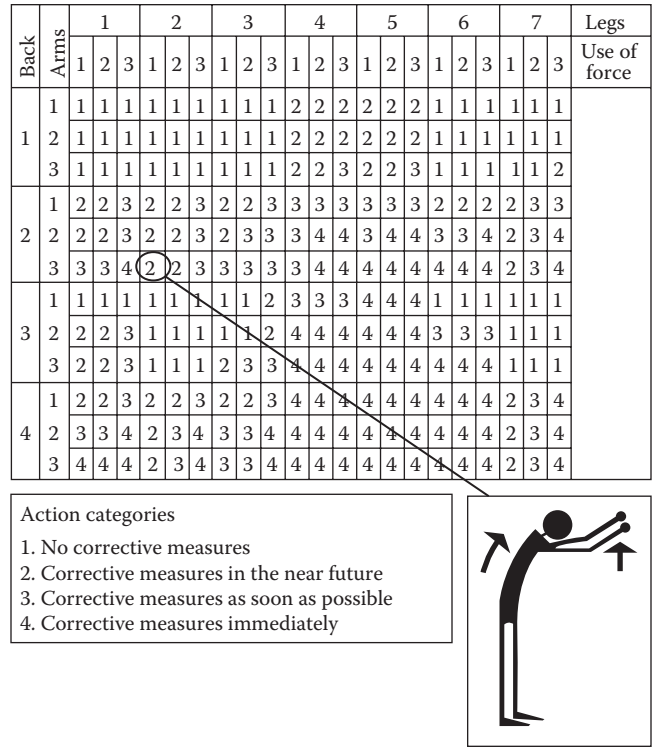


FIGURE 16.19 Action categories of the OWAS method for the task based on the posture and force codes.

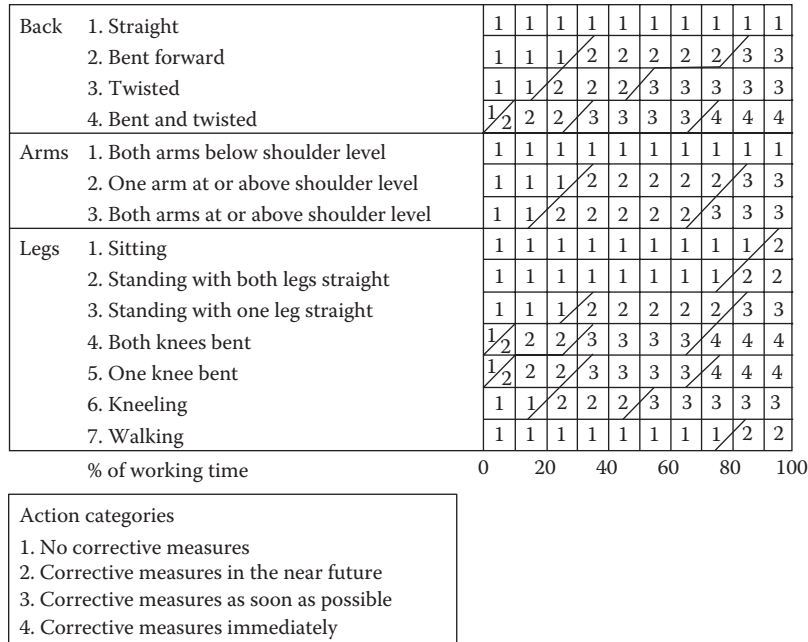
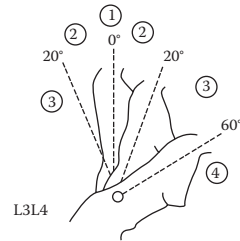


FIGURE 16.20 Action categories of the OWAS method for back, arm and leg postures determined from posture codes and percentage of working time.

## Group A

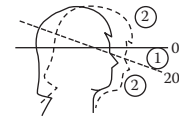
## Trunk

| Movement                           | Score | Change score:<br>+1 if twisting<br>or side flexed |
|------------------------------------|-------|---------------------------------------------------|
| Upright                            | 1     |                                                   |
| 0°–20° flexion<br>0°–20° extension | 2     |                                                   |
| 20°–60° flexion<br>>20° extension  | 3     |                                                   |
| >60° flexion                       | 4     |                                                   |



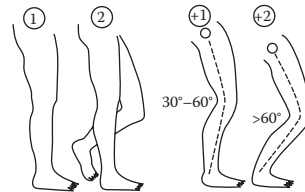
## Neck

| Movement                     | Score | Change score:<br>+1 if twisting<br>or side flexed |
|------------------------------|-------|---------------------------------------------------|
| 0°–20° flexion               | 1     |                                                   |
| >20° flexion or<br>extension | 2     |                                                   |



## Legs

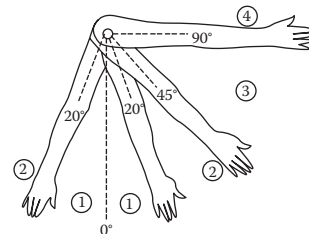
| Position                                                                       | Score | change score:<br>+1 if knee(s) between<br>30° and 60° flexion<br><br>+2 if knee(s) >60°<br>flexion<br>(N.B. not for sitting) |
|--------------------------------------------------------------------------------|-------|------------------------------------------------------------------------------------------------------------------------------|
| Bilateral weight bearing,<br>walking or sitting                                | 1     |                                                                                                                              |
| Unilateral weight-bearing,<br>feather weight-bearing or<br>an unstable posture | 2     |                                                                                                                              |



## Group B

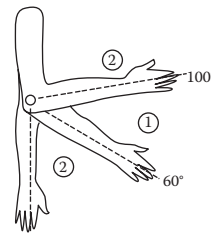
## Upper arms

| Position                          | Score | Change score:<br>+1 if arm is:<br>• Abducted<br>• Rotated<br><br>+1 if shoulder is raised<br><br>–1 if leaning, supporting<br>weight of arm or if<br>posture is gravity<br>assisted |
|-----------------------------------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 20° extension to<br>20° flexion   | 1     |                                                                                                                                                                                     |
| >20° extension<br>20°–45° flexion | 2     |                                                                                                                                                                                     |
| 45°–90° flexion                   | 3     |                                                                                                                                                                                     |
| >90° flexion                      | 4     |                                                                                                                                                                                     |



## Lower arms

| Movement                      | Score |
|-------------------------------|-------|
| 60°–100° flexion              | 1     |
| <60° flexion<br>>100° flexion | 2     |



## Wrists

| Movement                 | Score | Change score:<br>+1 if wrist is deviated<br>or twisted |
|--------------------------|-------|--------------------------------------------------------|
| 0°–15° flexion/extension | 1     |                                                        |
| >15° flexion/extension   | 2     |                                                        |

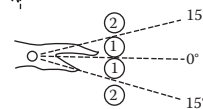


FIGURE 16.21 Posture scores for Groups A and B using the REBA method.

| Table A |  |      |   |   |   |   |   |   |   |   |   |   |   |
|---------|--|------|---|---|---|---|---|---|---|---|---|---|---|
| Trunk   |  | Neck |   |   |   |   |   |   |   |   |   |   |   |
|         |  | 1    |   |   |   | 2 |   |   |   | 3 |   |   |   |
|         |  | Legs |   |   |   |   |   |   |   |   |   |   |   |
| 1       |  | 1    | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 2       |  | 2    | 3 | 4 | 5 | 3 | 4 | 5 | 6 | 4 | 5 | 6 | 7 |
| 3       |  | 2    | 4 | 5 | 6 | 4 | 5 | 6 | 7 | 5 | 6 | 7 | 8 |
| 4       |  | 3    | 5 | 6 | 7 | 5 | 6 | 7 | 8 | 6 | 7 | 8 | 9 |
| 5       |  | 4    | 6 | 7 | 8 | 6 | 7 | 8 | 9 | 7 | 8 | 9 | 9 |

| Load/force |         |        |                                  |
|------------|---------|--------|----------------------------------|
| 0          | 1       | 2      | -1                               |
| <5 kg      | 5-10 kg | >10 kg | Shock or rapid build-up of force |

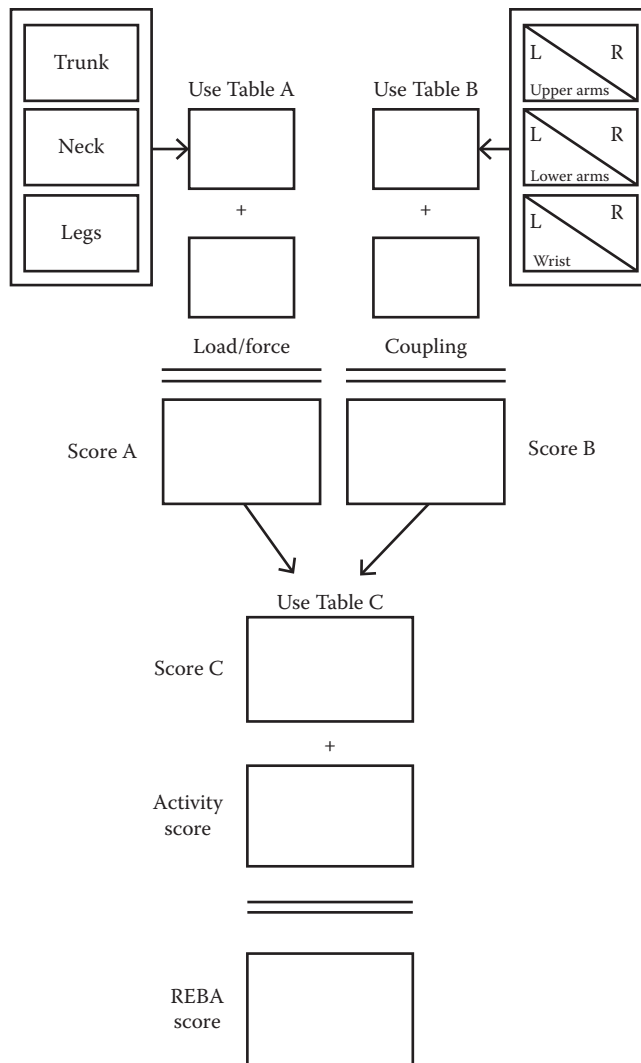
| Table B   |  |           |   |   |   |   |   |                                                 |                                                                                            |                                             |                                                                                                |  |
|-----------|--|-----------|---|---|---|---|---|-------------------------------------------------|--------------------------------------------------------------------------------------------|---------------------------------------------|------------------------------------------------------------------------------------------------|--|
| Upper arm |  | Lower arm |   |   |   |   |   | Coupling                                        |                                                                                            |                                             |                                                                                                |  |
|           |  | 1         |   |   | 2 |   |   | 0                                               | 1                                                                                          | 2                                           | 3                                                                                              |  |
|           |  | Wrist     |   |   |   |   |   | Good                                            | Fair                                                                                       | Poor                                        | Unacceptable                                                                                   |  |
| 1         |  | 1         | 2 | 2 | 1 | 2 | 3 | Well fitting handle and a mid-range power grip. | Hand hold acceptable but not ideal or coupling is acceptable via another part of the body. | Hand hold not acceptable although possible. | Awkward, unsafe grip, no handles<br><br>Coupling is unacceptable using other parts of the body |  |
| 2         |  | 1         | 2 | 3 | 2 | 3 | 4 |                                                 |                                                                                            |                                             |                                                                                                |  |
| 3         |  | 3         | 4 | 5 | 4 | 5 | 5 |                                                 |                                                                                            |                                             |                                                                                                |  |
| 4         |  | 4         | 5 | 5 | 5 | 6 | 7 |                                                 |                                                                                            |                                             |                                                                                                |  |
| 5         |  | 6         | 7 | 8 | 7 | 8 | 8 |                                                 |                                                                                            |                                             |                                                                                                |  |
| 6         |  | 7         | 8 | 8 | 8 | 9 | 9 |                                                 |                                                                                            |                                             |                                                                                                |  |

| Table C |         |    |    |    |    |    |    |    |    |    |    |    |    |
|---------|---------|----|----|----|----|----|----|----|----|----|----|----|----|
| Score A | Score B |    |    |    |    |    |    |    |    |    |    |    |    |
|         |         | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 |
|         | 1       | 1  | 1  | 1  | 2  | 3  | 3  | 4  | 5  | 6  | 7  | 7  | 7  |
|         | 2       | 1  | 2  | 2  | 3  | 4  | 4  | 5  | 6  | 6  | 7  | 7  | 8  |
|         | 3       | 2  | 3  | 3  | 3  | 4  | 5  | 6  | 7  | 7  | 8  | 8  | 8  |
|         | 4       | 3  | 4  | 4  | 4  | 5  | 6  | 7  | 8  | 8  | 9  | 9  | 9  |
|         | 5       | 4  | 4  | 4  | 5  | 6  | 7  | 8  | 8  | 9  | 9  | 9  | 9  |
|         | 6       | 6  | 6  | 6  | 7  | 8  | 8  | 9  | 9  | 10 | 10 | 10 | 10 |
|         | 7       | 7  | 7  | 7  | 8  | 9  | 9  | 9  | 10 | 10 | 11 | 11 | 11 |
|         | 8       | 8  | 8  | 8  | 9  | 10 | 10 | 10 | 10 | 10 | 11 | 11 | 11 |
|         | 9       | 9  | 9  | 9  | 10 | 10 | 10 | 11 | 11 | 11 | 12 | 12 | 12 |
|         | 10      | 10 | 10 | 10 | 11 | 11 | 11 | 11 | 12 | 12 | 12 | 12 | 12 |
|         | 11      | 11 | 11 | 11 | 11 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
|         | 12      | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |

| Activity score |                                                                                                     |
|----------------|-----------------------------------------------------------------------------------------------------|
| +1             | One or more body parts are static, e.g. held for longer than 1 min                                  |
| +1             | Action causes rapid large range changes in posture or an unstable base                              |
| +1             | Repeated small range actions, e.g. repeated more than four times per minute (not including walking) |

**FIGURE 16.22** Tables for the body (A), arm (B) and combined table (C) used to determine the final score using the REBA method.

are recorded on a REBA score sheet (Figure 16.23). Table A is used to determine the Group A score based on the posture scores of the trunk, neck and legs. A Group A force score is also determined from provided guidelines. Table B is used to determine the Group B posture and force scores. The posture and force scores for each group are added yielding Score A (for Group A) and Score B (for Group B). A third table, Table C, is used to determine Score C based on Scores A and B. Additionally, an activity score is determined from provided guidelines. Score C is then added to the activity score, yielding a final REBA score. The REBA score provides an indication of the relative action level, risk level and proposed actions (Table 16.14).



**FIGURE 16.23** Score sheet for the REBA method.

**TABLE 16.14**  
**Action Levels for the REBA Method**

| Action Level | REBA Score | Risk Level | Action<br>(Including Further Assessment) |
|--------------|------------|------------|------------------------------------------|
| 0            | 1          | Negligible | None necessary                           |
| 1            | 2–3        | Low        | May be necessary                         |
| 2            | 4–7        | Medium     | Necessary                                |
| 3            | 8–10       | High       | Necessary soon                           |
| 4            | 11–15      | Very high  | Necessary now                            |

## CONCLUSION

There are a variety of tools available for the practising ergonomist that allow for a systematic exposure assessment in workplaces without the high price tag of laboratory simulations or invasive measurement systems. The type of tool utilised will depend on the frequency of the task, the duration of the task, the risk factors present and quite possibly the biggest limitation will be the cost of the analysis tool. Some measurement techniques such as the MAC tool, Revised NIOSH Lifting Equation, RULA and REBA require only analyst hours and can be performed using a pencil and paper with minimal measurements. Other direct measurement tools such as pressure sensors, force gauges and motion capture systems require purchasing equipment that may not be economically viable if the analysis will be performed only once. It is up to the ergonomist to determine the most cost-effective and efficient method for performing their analysis. Due to the time and sometimes money invested in becoming proficient in a particular measurement or analysis method, the individual ergonomist often develops their own repertoire of tools and methods. The broad range of methods presented in this chapter offers considerable variety in terms of technical sophistication, cost and purpose for addressing most of the physical ergonomics issues encountered in the workplace.

## DISCLAIMER

The findings and conclusions in this chapter are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health (NIOSH). Mention of any product does not constitute an endorsement by the NIOSH.

## REFERENCES

- Andersson, K., Karlehagen, S. and Jonsson, B. (1987). The importance of variations in questionnaire administration. *Applied Ergonomics*, **18**(3), 229–232.
- Bao, S. and Silverstein, B. (2005). Estimation of hand force in ergonomic job evaluations. *Ergonomics*, **48**(3), 288–301.
- Bernard B. (1997). *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-Related Musculoskeletal Disorders for the Neck, Upper Extremity, and Low Back*. Cincinnati, OH: Department of Health and Human Services, National Institute for Occupational Safety and Health.
- Borenstein, J., Ojeda, L. and Kwanmuang, S. (2009). Heuristic reduction of gyro drift in IMU-based personnel tracking systems. Presented at the *SPIE Defense, Security + Sensing, Conference 7306A: Optics and Photonics in Global Homeland Security V*, Orlando, FL, 13–17 April 2009.
- Brodie, M., Walmsley, A. and Page, W. (2008). Fusion motion capture: A prototype system using inertial measurement units and GPS for the biomechanical analysis of ski racing. *Sports Technology*, **1**(1), 17–28.
- Burden, A. and Bartlett, R. (1999). Normalisation of EMG amplitude: An evaluation and comparison of old and new methods. *Medical Engineering & Physics*, **21**, 247–257.
- Casey, J.S., McGorry, R.W. and Dempsey, P.G. (October 2002). Getting a grip on grip force estimates. *Professional Safety*, 18–24.
- Cavallo, F., Sabatini, A.M. and Genovese, V. (2005). A step toward GPS/INS personal navigation systems: Real-time assessment of gait by foot inertial sensing. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Diego, CA.
- Chaffin, D.B., Andersson, G.J. and Bernard, B.J. (1999). *Occupational Biomechanics* (3rd edn.), New York: Wiley Interscience.
- Chaffin, D.B., Lee, M. and Freivalds, A. (1980). Muscle strength assessment from EMG analysis. *Medicine and Science in Sports and Exercise*, **12**(3), 205–211.
- Chang, J.H., Wu, J.D., Liu, C.Y. and Hsu, D.J. (2012). Prevalence of musculoskeletal disorders and ergonomic assessments of cleaners. *American Journal of Industrial Medicine*, **55**(7), 593–604.
- Corlett, E.N. and Bishop, R.P. (1976). A technique for assessing postural discomfort. *Ergonomics*, **19**, 175–182.
- Dempsey, P.G. (2003). A survey of lifting and lowering tasks. *International Journal of Industrial Ergonomics*, **33**(1), 11–16.



- Dempsey, P.G. (2006). Psychophysical approach to task analysis. In: *The Occupational Ergonomics Handbook: Fundamentals and Assessment Tools for Occupational Ergonomics* (2nd edn.), W.S. Marras and W. Karwowski (eds.), pp. 47.1–47.31. Boca Raton, FL: CRC Press.
- Dempsey, P.G., Ciriello, V.M., Maikala, R.V. and O'Brien, N.V. (2008). Oxygen consumption prediction models for individual and combination materials handling tasks. *Ergonomics*, **51**(11), 1776–1789.
- Dempsey, P.G. and Leamon, T.B. (1999). Model ergonomics assessment: Retail ice cream shop case study. In: *Applied Ergonomics: Case Studies*, D.C. Alexander (ed.), Vol. 1. Norcross, GA: Engineering & Management Press.
- Dempsey, P.G. and Maynard, W.S. (2005). Manual materials handling: Using the liberty mutual tables to evaluate these tasks. *Professional Safety*, **50**(5), 20–25.
- Dempsey, P.G., McGorry, R.W., Cotnam, J. and Braun, T.W. (2000). Ergonomics investigation of retail ice cream operations. *Applied Ergonomics*, **31**(2), 121–130.
- Dempsey, P.G., McGorry, R.W. and Maynard, W.S. 2005. A survey of tools and methods used by certified professional ergonomists. *Applied Ergonomics*, **36**, 489–503.
- Dickinson, C.E., Campion, K., Foster, A.F., Newman, S.J., O'Rourke, A.M.T. and Thomas, P.G. (1992). Questionnaire development: An examination of the Nordic Musculoskeletal questionnaire. *Applied Ergonomics*, **23**(3), 197–201.
- Dolan, P. and Adams, M.A. (1993). The relationship between EMG activity and extensor moment generation in the erector spinae muscles during bending and lifting activities. *Journal of Biomechanics*, **26**, 513–522.
- Drury, C.G. (1987). A biomechanical evaluation of the repetitive motion injury potential of industrial jobs. *Seminars in Occupational Medicine*, **2**(1), 41–50 (New York: Thieme Medical Publishers Inc.).
- Favre, J., Aissaoui, R., Jolles, B.M., de Guise, J.A. and Aminian, K. (2004). Functional calibration procedure for 3D knee joint angle description for inertial sensors. *Journal of Biomechanics*, **42**(4), 2330–2335.
- Favre, J., Jolles, B.M., Aissaoui, R. and Aminian, K. (2008). Ambulatory measurement of 3D knee joint angle. *Journal of Biomechanics*, **41**(5), 1029–1035.
- Finnish Institute of Occupational Health. (1992). *OWAS: A Method for the Evaluation of Postural Load during Work*. Helsinki, Finland: Finnish Institute of Occupational Health.
- Garg, A., Chaffin, D.B. and Herrin, G.D. (1978). Prediction of metabolic rates for manual materials handling jobs. *American Industrial Hygiene Association Journal*, **39**, 661–674.
- Hamberg-van Reenen, H.H., van der Beek, A.J., Blatter, B.M., van der Grinten, M.P., van Mechelen, W. and Bongers, P.M. (2008). Does musculoskeletal discomfort at work predict future musculoskeletal pain? *Ergonomics*, **51**(5), 637–648.
- Hanlon, M. and Anderson, R. (2009). Real-time gait event detection using wearable sensors. *Gait & Posture*, **30**, 523–527.
- Health Council of the Netherlands. (1995). Risk assessment of manual lifting (Pub. No. 1995/02). The Hague, the Netherlands: Health Council of the Netherlands.
- Hignett, S. and McAtamney, L. (2000). Rapid entire body assessment (REBA). *Applied Ergonomics*, **31**(2), 201–205.
- HSE. (2003). *Manual Handling Assessment Charts (The MAC)*. Sudbury, Ontario, Canada: HSE.
- HSE. (2004). *Manual Handling: Manual Handling Operations Regulations 1992 (as Amended), Guidance on Regulations L23* (3rd edn.). Sudbury, Ontario, Canada: HSE.
- HSE. (2010). *Assessment of Repetitive Tasks of the Upper Limbs (the ART Tool)*. Sudbury, Ontario, Canada: HSE.
- Hsiao, H., Guan, J. and Weatherly, M. (2002). Accuracy and precision of two in-shoe pressure measurement systems. *Ergonomics*, **45**(8), 537–555.
- Hsu, D.J., Chang, J.H., Wu, J.D., Chen, C.Y. and Yang, Y.H. (2011). Prevalence of musculoskeletal disorders and job exposure in Taiwan oyster shuckers. *American Journal of Industrial Medicine*, **54**(11), 885–893.
- Kose, A., Cereatti, A., Laudani, L. and Della Croce, U. (2011). Estimation of stride length in walking using a single inertial measurement unit attached to the waist. *Gait & Posture*, **33**(Suppl. 1), S49–S50.
- Kuorinka, I. and Forcier, L. (1995). *Work Related Musculoskeletal Disorders: A Reference Book for Prevention*. London, U.K.: Taylor & Francis Group.
- Kuorinka, I., Jonsson, B., Kilbom, A., Vinterberg, H. et al. (1987). Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Applied Ergonomics*, **18**(3), 233–237.
- Li, Q., Young, M., Naing, V. and Donelan, J.M. (2010). Walking speed estimation using shank-mounted inertial measurement unit. *Journal of Biomechanics*, **43**(8), 1640–1643.
- Lloyd, D.G. and Besier, T.F. (2003). An EMG-driven musculoskeletal model to estimate muscle forces and knee joint moments in vivo. *Journal of Biomechanics*, **36**(6), 765–776.

- Marras, W.S., Fathallah, F.A., Miller, R.J., Davis, S.W. and Mirka, G.A. (1992). Accuracy of a three-dimensional lumbar motion monitor for recording dynamic trunk motion characteristics. *International Journal of Industrial Ergonomics*, **9**(1), 75–87.
- Mazzà, C., Donati, D., McCamley, J., Picerno, P. and Cappozzo, A. (2012). An optimized Kalman filter for the estimate of trunk orientation from inertial sensors data during treadmill walking. *Gait & Posture*, **35**(1), 138–142.
- McAtamney, L. and Corlett, E.N. (1993). RULA: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, **24**(2), 91–99.
- McGorry, R.W. (2001). A system for the measurement of grip forces and applied moments during hand tool use. *Applied Ergonomics*, **32**, 271–279.
- Milner-Brown, H.S. and Stein, R.B. (1975). The relation between the surface electromyogram and muscular force. *Journal of Physiology*, **246**, 549–569.
- Mital, A., Nicholson, A.S. and Ayoub, M.M. (1997). *A Guide to Manual Materials Handling* (2nd edn.), London, U.K.: Taylor & Francis Group.
- NIOSH. (1981). Work practices guide for manual lifting, Pub. No. 81-122. Cincinnati, OH: DHHS(NIOSH).
- Nordstrom, D.L., Vierkant, R.A., DeStefano, F. and Layde, P.M. (1997). Risk factors for carpal tunnel syndrome in a general population. *Occupational and Environmental Medicine*, **54**, 734–740.
- Ojeda, L. and Borenstein, J. (2007). Personal dead-reckoning system for GPS-denied environments. *IEEE International Workshop on Safety, Security, and Rescue Robotics (SSRR2007)*, Rome, Italy, 27–29 September 2007.
- Paquet, V.L., Mathiassen, S.E. and Dempsey, P.G. (2006). Video-based ergonomic job analysis: A practitioner's guide. *Professional Safety*, **51**(11), 27–35.
- Perry, J. and Bekey, G.A. (1981). EMG-force relationships in skeletal muscle. *CRC Critical Reviews in Biomechanical Engineer*, **7**(1), 287–299.
- Pinder, A.D.J. and Frost, G.A. (2011). Prospective evaluation of the 1991 NIOSH lifting equation (Research Report RR901). London, U.K.: Health and Safety Executive.
- Potvin, J.R. (1997). Use of NIOSH equation inputs to calculate lumbosacral compression forces. *Ergonomics*, **40**(7), 691–707.
- Punnett, L. and Wegman, D.H. (2004). Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. *Journal of Electromyography and Kinesiology*, **14**(1), 13–23.
- Rempel, D.M. and Punnett, L. (1997). Epidemiology of wrist and hand disorders. In: *Musculoskeletal Disorders in the Workplace: Principles and Practice*, M. Nordin, G.B. Andersson and M.H. Pope (eds.). Philadelphia, PA: Mosby-Year Book, Inc., pp. 421–430.
- Snook, S.H. and Ciriello, V.M. (1991). The design of manual handling tasks: Revised tables of maximum acceptable weights and forces. *Ergonomics*, **34**, 1197–1213.
- van Dieën, J.H., Faber, G.S., Loos, R.C.C., Kuijer, P.P.F.M., Kingma, I., Molen, H.F. van der and Frings-Dresen, M.H.W. (2010). Validity of estimates of spinal compression forces obtained from worksite measurements. *Ergonomics*, **53**(6), 792–800.
- Waters, T.R., Putz-Anderson, V. and Garg, A. (1994). Applications manual for the revised NIOSH lifting equation, Pub. No. 94-110. Cincinnati, OH: DHHS(NIOSH).
- Waters, T.R., Putz-Anderson, V., Garg, A. and Fine, L.J. (1993). Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, **36**, 749–776.
- Werner, R.A., Franzblau, A., Albers, J.W., Buchele, H. and Armstrong, T.J. (1997). Use of screening nerve conduction studies for predicting future carpal tunnel syndrome. *Occupational and Environmental Medicine*, **54**, 96–100.
- Widanarko, B., Legg, S., Stevenson, M., Devereux, J., Eng, A., 't Mannetje, A., Cheng, S. and Pearce, N. (2012). Prevalence and work-related risk factors for reduced activities and absenteeism due to low back symptoms. *Applied Ergonomics*, **43**(4), 727–737.
- Wong, W.Y. and Wong, M.S. (2006). Using inertial sensing units to detect spinal posture change in trunk movements. *Gait & Posture*, **24**(2), S221–S222.
- Woods, J.J. and Bigland-Ritchie, B. (1983). Linear and non-linear EMG/force relationships in human muscles. *American Journal of Physical Medicine*, **62**(6), 287–299.

---

# 17 Biomechanical Basis of Ergonomics Assessment

*Patrick G. Dempsey and Moira Munro*

## CONTENTS

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| Introduction.....                                               | 488 |
| Relevant Tasks.....                                             | 488 |
| Interpretation of Results.....                                  | 488 |
| How Detailed Should the Analysis Be? .....                      | 489 |
| Equipment Required.....                                         | 489 |
| Human Variability .....                                         | 490 |
| Principles of Biomechanical Calculations .....                  | 490 |
| Moments and Lever Arms.....                                     | 490 |
| A Simple Example with a 2D Low-Back Model .....                 | 491 |
| Moments in 3D Space .....                                       | 492 |
| Moments in Dynamic Tasks.....                                   | 496 |
| Prediction of Strength and Task Feasibility Using Moments ..... | 497 |
| Reliability of Maximum Joint Strength Data.....                 | 497 |
| Dependence of Joint Strength with Joint Angles .....            | 498 |
| Interpretation of Results.....                                  | 498 |
| Forces on the Low Back.....                                     | 500 |
| Guidelines from Low-Back Forces .....                           | 500 |
| Applications of Low-Back Modelling.....                         | 500 |
| Models of the Low Back.....                                     | 501 |
| A Simple 2D Model .....                                         | 501 |
| Three-Dimensional Models.....                                   | 503 |
| 3D Models Requiring Optimisation .....                          | 505 |
| Three-Dimensional Models Requiring Biological Input.....        | 507 |
| Inputs to Biomechanical Calculations .....                      | 507 |
| Posture Input .....                                             | 507 |
| Body Segment Weights .....                                      | 508 |
| Low-Back Geometry .....                                         | 508 |
| Role of Intra-Abdominal Pressure .....                          | 508 |
| Angle of Discs.....                                             | 510 |
| Uses and Limitations of Biomechanics.....                       | 511 |
| Repetitive Work and Fatigue .....                               | 511 |
| Sources of Inaccuracy .....                                     | 511 |
| Conclusions.....                                                | 512 |
| Acknowledgement .....                                           | 512 |
| Disclaimer.....                                                 | 512 |
| References.....                                                 | 512 |

## INTRODUCTION

Biomechanics is the study of forces generated by and acting on the human body. This field of study is used in the ergonomics/human factors (E/HF) field most often to assess manual handling tasks and other tasks that require generating forces both internal and external to the body. This chapter outlines the types of task for which a biomechanical analysis is or is not an appropriate method, details an overview of the tools required to carry out an analysis and describes various methods and models used in the field along with their advantages and disadvantages. This chapter focuses on the fundamental concepts and calculations that are required to perform a biomechanical analysis as biomechanical models and the requisite equipment are constantly evolving.

When interpreting results of a biomechanical analysis, we come to the questions of (1) how valid are the simplifying assumptions and approximations that are made in the analysis, (2) how robust are the results of the analysis and (3) what are our criteria for 'safe' levels of force and how reliable are they? Just as the approach used to estimate forces on the body is important for a reasonably valid estimate, the interpretation of these results is also critical.

## RELEVANT TASKS

Biomechanics is a useful tool in most manual handling situations, whether people are lifting, lowering, pushing, pulling or even when no load is handled but the body's own weight is creating postural stress. Powered and non-powered tool use is another type of work suited for biomechanical analysis. In general, however, applications involving tools are often focused on high-force activities requiring gross body motions.

Biomechanics is best used as a comparative method because, as will be shown throughout this chapter, all biomechanical modelling results rest on simplifications and approximations as direct measurement of forces acting on and within living body tissues is not possible, except in rare experimental settings. Because of these factors, biomechanics is particularly successful at demonstrating possible improvements obtained from redesigning a task or helping the ergonomist choose between alternate tasks or workplace designs. Biomechanics can also be used to identify the most stressful tasks or elements of a job so that resources can be directed to the most harmful aspects first.

## INTERPRETATION OF RESULTS

Different tasks or workplace designs can be compared with calculated forces at joints such as elbows or shoulders and for forces on the spine or within trunk muscles. It is also possible to assess the feasibility and estimate the safety of tasks even at the design stage by comparing these forces with recommended limits or strength data in the literature. Unfortunately, we are not at the stage of answering precisely 'how much is too much' for safe levels of exertion and force that do not lead to injury.

Drury et al.'s (1983) evaluation of a palletising aid provides a good example of the use of biomechanics with other methods. Very simple calculations based on video recordings showed to what extent the load on the lumbar spine was reduced when a palletising aid was used. The heart rate was also shown to go down, and the authors used these factors and injury statistics to evaluate the cost effectiveness of introducing the aids. If this study had been based only on accepted 'safe' spinal load limits, this particular biomechanical model would have indicated that they were quite acceptable. The use of several methods to evaluate the task ensured that problems were not overlooked and that biomechanics demonstrated the benefits of introducing an aid.

## HOW DETAILED SHOULD THE ANALYSIS BE?

The simplest biomechanical calculations are those relating to a static posture and force in the sagittal plane (with no twisting or lateral bending). The problem is purely 2D, and there is no consideration of forces caused by accelerations and inertia. It is easy to record the posture in that one plane, and all the calculations can be done with a calculator.

If the task is characterised by much lateral bending or twisting, problems or possible improvements would be overlooked with a 2D analysis. Recording posture in three dimensions requires special methods outlined later. Calculations are quite lengthy and a computer program is necessary. A 3D analysis is therefore quite time-consuming unless the posture-recording method and computer program have already been set up.

If the task is not static, the forces resulting from the accelerations need to be included in the calculations. This requires sampling the posture at a reasonable frequency (e.g. 50 Hz or higher), along with the value and direction of the acceleration of each limb. Dynamic analyses were once restricted to the laboratory, although technological advances such as less-expensive video-based systems, gaming technology and inertial sensors are increasing the practicality of conducting dynamic analyses in the workplace (e.g. Dutta, 2012; Thewlis et al., 2013; Vignais et al., 2013). In 2D it is possible sometimes to avoid recording posture and acceleration continuously by using simulations. This has been done for lifting with the accuracy of the results depending on how similar the lifting technique is to the one used as a database by the computer program. Finally, it is quite common to analyse dynamic tasks statically, freezing at the beginning, middle and end of the task, for instance. If the task was done slowly and smoothly, this is a good approximation. Otherwise, this may underestimate forces up to two or three times during the acceleration phase (Garg et al., 1982). These approaches will be discussed in more detail later.

## EQUIPMENT REQUIRED

A biomechanical analysis requires measurements of posture, hand force (and any other external forces acting on the body) and, in the case of dynamic tasks, accelerations. Hand forces can be measured with equipment as simple as a spring balance, but in some cases a tool equipped with strain gauges is required. These tools can be interfaced to a computer for online recordings. For pushing and pulling, analogue and digital scales are available commercially, including load cells designed for generic applications. These load cells can be connected to personal computers or smaller portable electronic devices (McGorry et al., 2002). Additional methods are discussed in Chapter 16.

The choice of a posture recording method will depend on whether the analysis is 2D or 3D and in the field or in the laboratory. It may not be necessary to record posture when using some 2D computer programs because the posture is input via a stickman on the screen or can be estimated from a photograph. It is best to base the posture input on a photograph in order to input an accurate posture. Videos can be digitised for 2D or 3D posture recordings; this is a time-consuming process in 3D but is often used as a simple field method in 2D.

For 3D analyses of dynamic tasks, a dedicated acquisition system is often required. A comprehensive review goes beyond the scope of this chapter, but there are numerous systems available on the market that collect postural data at rates sufficient for dynamic analyses. These systems use various approaches including retrieving the postural data from standard video, tracking active and passive motion markers, and tracking postures using electromagnetic fields. Most systems are costly and are typically purchased by those with ongoing research or a significant amount of field analysis. More recently, less-expensive systems and technology from other applications such as gaming (e.g. Dutta, 2012) have been investigated as alternatives to sophisticated systems that are difficult to use in workplaces. The proliferation of cost-effective wearable inertial sensors and measurement units has also led to an increase in investigations of their use for biomechanics (e.g. Fong and Chan, 2010;

Vignais et al., 2013). Both types of systems minimise the need for additional equipment in the working area that can interfere with workers or equipment.

## HUMAN VARIABILITY

Biomechanics is sometimes used with the aim of determining safe force or torque limits for a high percentage of the population. At this stage, human variability must be borne in mind. Two critical points are the large ranges in muscle strength in the population and the strength of the spinal column. Predicting the percentage of the population with the joint strength capable of performing a task and developing criteria for safe limits based on compression tolerance of the spine are two of the most common applications of biomechanics in E/HF.

Human variability in body weight and stature also affects biomechanical calculations. The posture that people adopt for a given task affects loading, but given a particular posture, the loads on the body are greater for higher body weights and longer limb lengths. It is therefore recommended to use the 95th percentile body weight and stature when using biomechanics as a basis for workplace or task design in order to err on the side of safety.

Finally, when results of biomechanical calculations are evaluated against 'safe' limits, it must be borne in mind that people differ in their susceptibility to back pain or injury. Hutton and Adams (1982) have shown that, as a rule, female spines are more susceptible to fracture than male spines and that age is also a weakening factor. Troup et al. (1987) found that people who had experienced back pain chose to lift lighter loads. Ferguson et al. (2005) provided recommendations for workers returning to work following a low-back pain episode that included reduced lifting limits compared to asymptomatic workers. The mechanism of back pain is still uncertain in many cases; tests on cadaveric spines do not necessarily produce the same effects as those observed in vivo (Brinckmann, 1986).

## PRINCIPLES OF BIOMECHANICAL CALCULATIONS

This section demonstrates how loads at body joints are estimated. The results of these estimates can be used to compare alternate task designs or they can be compared with population data on maximum strength capabilities in order to assess the strenuousness of the task.\* This section will also demonstrate how forces within the low back can be evaluated with a simple 2D model.

The calculation of forces is derived from Newton's third law of motion that all forces must balance each other if the body is to be in equilibrium. If there is a resultant force in any direction, the body will move in that direction. The calculation of loads at body joints discussed next is in fact the calculation of moments or turning forces around a point. Moments must also balance so that the sum of moments around any point is zero if there is to be no rotation.

## MOMENTS AND LEVER ARMS

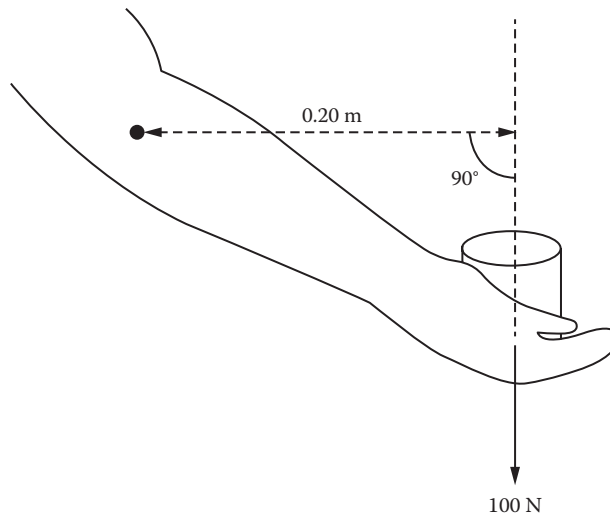
The moment, or torque, of a force around a point is a measure of the turning force around the point. For example, holding a weight in the hand creates a moment about the elbow and tends to make the elbow extend. Muscles spanning the elbow provide the opposite moment by contracting so that the elbow is able to support the weight.

The moment of a force about a point is the product of the force and the perpendicular distance between the point and the line of action of the force (see Figure 17.1). A weight of 100 N<sup>†</sup> held in the

\* This last aspect is discussed in the next main section. Readers who do not need to know the details of calculations can turn directly to that section after reading the section 'Moments and Lever Arms'.

<sup>†</sup> Newtons (N) are units of force or weight. A 1 kg mass weighs 9.81 N. More on this and other units appears at the end of this section.





**FIGURE 17.1** Moment at the elbow due to the 100 N weight held at the hand is the product of the force (100 N) and perpendicular distance (0.20 m) through which it acts.

hand creates a moment of  $100 \times 0.20 = 20 \text{ N m}$  (Newton metres) for the position shown. If the weight was held with the arm hanging down, the lever arm of the force would be zero, and so the moment about the elbow would be zero. There would, of course, still be a force at the elbow, resisting the downward pull of the 100 N weight.

Another force exerting a moment about the elbow is the weight of the hand and forearm. The location of the centre of gravity of the hand and forearm, and their weight, can be estimated from the tables presented later in this chapter (see section 'Inputs to Biomechanical Calculations').

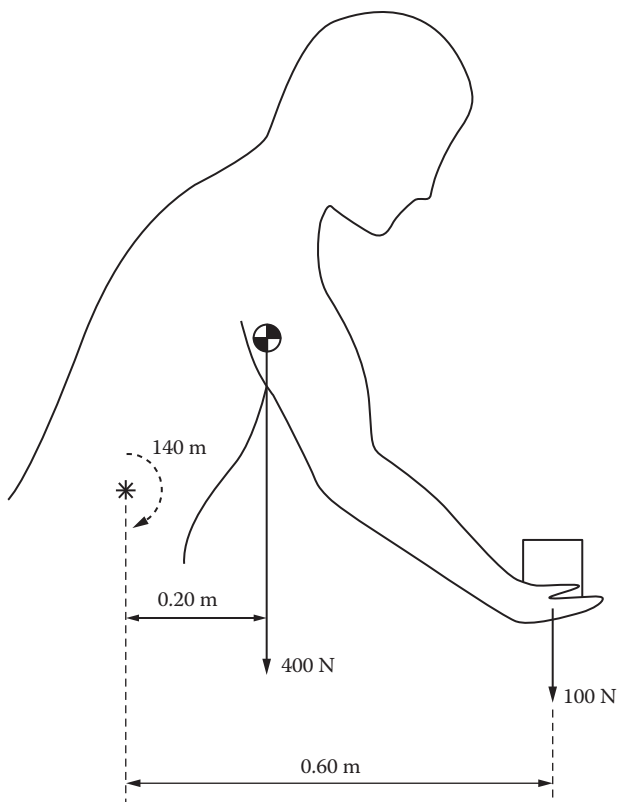
### A SIMPLE EXAMPLE WITH A 2D LOW-BACK MODEL

To evaluate forces in the lumbar region, the moment around a point of the low back is calculated in the same way as has been demonstrated for the elbow. A model of the muscle geometry is then used to calculate how much force the back muscles exert to counteract this moment. This enables the compression force on the spine to be estimated, which is a commonly used criterion.

The result is a simple example in 2D (see Figure 17.2) with the posture and hand force in the sagittal plane. Calculations in 3D, for lateral bending and twisting, will follow after that.

A participant is depicted holding a 100 N weight, and moments are calculated around the point indicated by a star, which is situated on a point of the lumbar spine. The weight of the body above this point is 400 N in this example, acting with a lever arm of 0.20 m and creating an 80 N m moment. Add to this the effect of the weight at the hands, acting with a 0.60 m lever arm, and the total moment created around the starred point is 140 N m. Thus, the weight at the hands and the participant's own body weight tend to flex the trunk; trunk muscles and ligaments must counteract this so that the posture is held.

Forces within the trunk can be evaluated at this stage using a model of the low back. Such models vary in detail and complexity, as will be seen later, but the principle can be shown here. In a simple 2D model, the 140 N m trunk flexion moment is resisted by back muscles alone. The greater the leverage those muscles have from the spine, the smaller the force needed from them. Let us suppose the line of action of the back muscles is 5.8 cm posterior to the spine (this choice of values will be discussed later). The force these muscles need to exert is  $140/0.058 = 2414 \text{ N}$ . This is much larger than the body weight or the hand force because the muscles are balancing the moments through a very small lever arm.



**FIGURE 17.2** Forces and low-back moment 100 N at the hands.

As the back muscles pull to counteract moments, they compress the lumbar spine. The weight of the body above the lumbar spine and the weight at the hands also compress it, so finally the total compression force is the sum of all these components:  $100 + 400 + 2414 = 2914$  N. This again may seem high, but compression tests have shown that the spine can, in general, withstand this type of force. Further discussion on this is given in section 'Forces on the Low Back'.

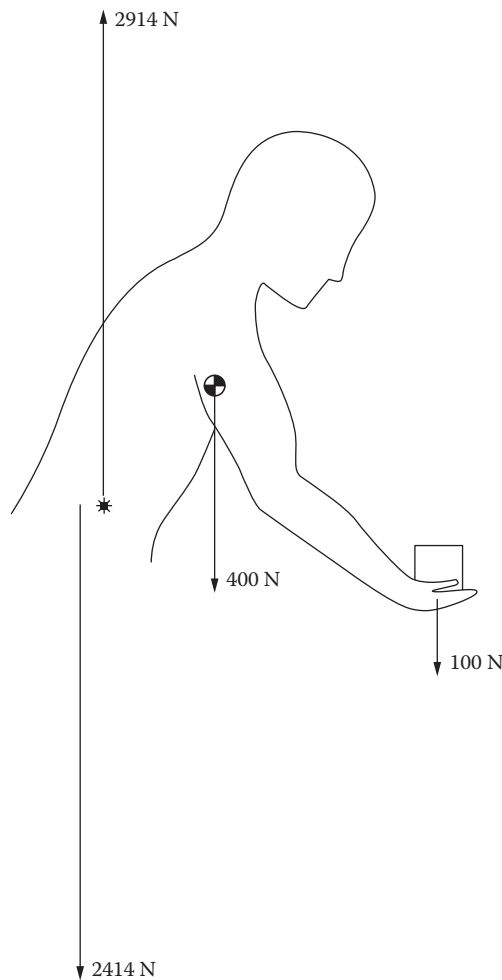
Figure 17.3 shows all the forces involved in this simple problem – the sum of all these forces is zero, and so is the sum of all moments around any point. The point indicated by a star on the lumbar spine was chosen only to simplify calculations as the compression force does not create a moment about this point. In this example, the centre of mass of the upper part of the body is shown as being 0.20 m from the lumbar spine. In fact, this would not be known in an ordinary problem – what is estimated is the location of the centre of mass of each body segment, and so the moments created by each segment can be summed.\* We will now go on to calculations in 3D static or dynamic tasks.

### MOMENTS IN 3D SPACE

This section, on moments in 3D space, contains details which do not need to be read by users who will not be actually performing calculations.

Previous examples were restricted to postures and forces in the sagittal plane. For asymmetric postures or forces not contained in this plane, the calculation of the moments can be made

\* The section 'Moments in 3D Space' shows in detail how this is done. Calculations are simplified because in 2D situations  $r_y = 0$ ,  $F_y = 0$ ,  $M_x = 0$ ,  $M_z = 0$ .

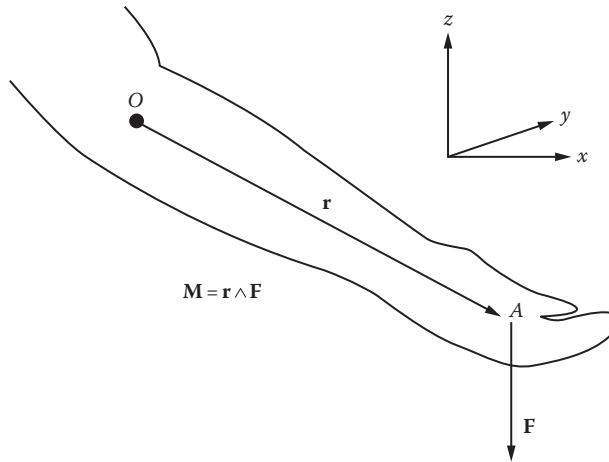


**FIGURE 17.3** Equilibrium of forces when holding a 100 N weight.

by the following method. Figure 17.4 shows a force **F** (bold type indicates a vector) acting at the hand, with the vector **r** running from the elbow to the hand. The simplest way to determine the moment, **M**, around the elbow is to record the  $x$ ,  $y$ ,  $z$  components of vectors **r** and **F** along a set of perpendicular axes. Most measuring equipment will allow this. Thus, the components of **r** are  $r_x$ ,  $r_y$  and  $r_z$  along the  $x$ ,  $y$  and  $z$  axes, and those of **F** are  $F_x$ ,  $F_y$  and  $F_z$ . The moment **M** around the elbow is also a vector, and it is simple to calculate its components  $M_x$ ,  $M_y$  and  $M_z$ ,  $M_x$  is the turning force in the ( $y$ ,  $z$ ) plane,  $M_y$  is the turning force in the ( $z$ ,  $x$ ) plane and  $M_z$  is the turning force in the ( $x$ ,  $y$ ) plane.

The resultant of  $M_x$ ,  $M_y$  and  $M_z$  is the size of the vector **M**. From Pythagoras' theorem,

$$M^2 = M_x^2 + M_y^2 + M_z^2$$



**FIGURE 17.4** Moment of a force about  $O$  is the vector product (noted  $\wedge$ ) of the vector  $\mathbf{r}$  running from  $O$  to  $A$ , with the force vector  $\mathbf{F}$ .

$M_x$ ,  $M_y$  and  $M_z$  are obtained through the following equations:

$$M_x = r_y F_z - r_z F_y$$

$$M_y = r_z F_x - r_x F_z$$

$$M_z = r_x F_y - r_y F_x$$

A shorthand notation for these is

$$\mathbf{M} = \mathbf{r} \wedge \mathbf{F}$$

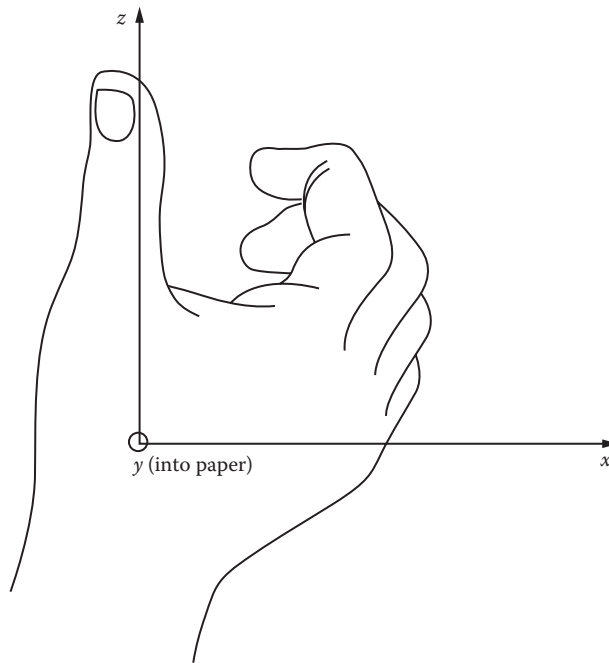
where  $\mathbf{M}$  is described as the vector product (noted  $\wedge$ ) of  $\mathbf{r}$  and  $\mathbf{F}$  (the order is important).

For calculations of moments in a 2D situation, we have seen that the moment was the product of the force and its perpendicular distance (lever arm). This gives the same result as the aforementioned equations, so the method chosen depends on which is easiest to record: the lever arm or the components along the  $x$ ,  $y$  and  $z$  axes.

Each component of the moments,  $M_x$ ,  $M_y$  and  $M_z$ , represents the turning force about the  $x$ ,  $y$  or  $z$  axis. For instance, in Figure 17.4, the  $y$  axis is directed into the paper, and  $M_y$  is the moment about that axis and represents the flexion/extension moment about the elbow.

This is the only moment in the case of Figure 17.4; the reader can verify from the aforementioned equations that  $M_x$  and  $M_z$  are zero. (The only component of  $\mathbf{F}$  is along the  $z$  axis, and  $\mathbf{r}$  and  $\mathbf{F}$  are in one plane:  $r_y = 0$ ,  $F_x = 0$ ,  $F_y = 0$ .) This means that there are no twisting or abduction/adduction requirements on the elbow.

The sign (positive or negative) of a moment indicates the direction of the turning force. In the example of Figure 17.4,  $F_z$  is negative, so from the aforementioned equations,  $M_y$  is positive. This represents the extension effect force  $\mathbf{F}$  has on the elbow. A negative  $M_y$  would represent a flexion effect.



**FIGURE 17.5** Conventions for a system of axes.

One method of working out the meaning of a positive or negative moment is as follows: stick out your right-hand thumb in the direction of the selected axis,  $y$  in this case. Your other fingers naturally curl round in the direction of the rotation corresponding to a positive  $M_y$ , in this case, elbow extension.

For this system, a right-handed set of axes is needed, that is,  $y$  should go into the paper, as in Figure 17.4, not out of it. An easy trick to ensure the axes are right-handed is to point the right-hand thumb along the  $z$  axis. The fingers curl around to indicate the direction from  $x$  to  $y$  (Figure 17.5).

So far we have looked at the moment in 3D created by one force,  $\mathbf{F}$ . When calculating the moment about a point in the low back, for instance, the weights of several body segments have to be taken into account. One method is to add up the moments that each of these forces creates about the low back. This is summarised by the following equation:

$$\mathbf{M} = \mathbf{r} \wedge \mathbf{F} + \mathbf{r}_1 \wedge m_1 \mathbf{g} + \mathbf{r}_2 \wedge m_2 \mathbf{g} + \dots$$

where

$\mathbf{F}$  is an external force acting on the body, such as a weight at the hands

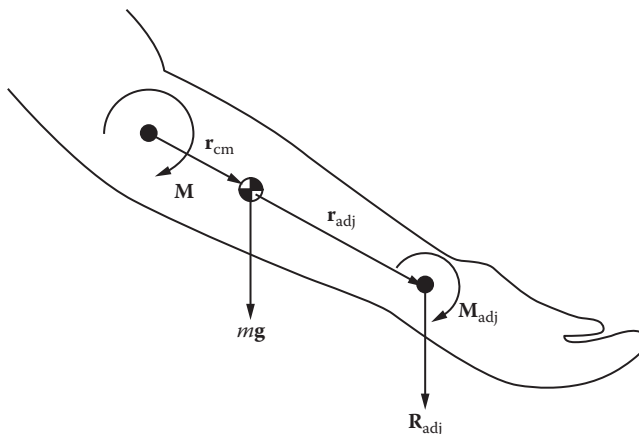
$\mathbf{r}$  is the vector running from the low back to the point of application of  $\mathbf{F}$

$m_1$  and  $m_2$  are the masses of body segments (kg)

$\mathbf{r}_1$  and  $\mathbf{r}_2$  are the vectors running from the low back to the centres of mass of body segments

$\mathbf{g}$  is the acceleration due to gravity ( $9.81 \text{ m/s}^2$ , downwards)

A second method to calculate  $\mathbf{M}$ , which is useful when moments at several joints of the body are already known, is to calculate the moment around the wrist, then use this result and add the effect of the forearm weight to work up to the elbow, and so on to the shoulder, until the low back is reached.



**FIGURE 17.6** Moment  $M_{adj}$  at the wrist can be used to calculate the moment  $M$  at the elbow.

This second method, although it may seem less immediate, is more economical if the moments at the wrist, elbow and so on were required anyway. The following equation is then used (symbols are shown in Figure 17.6):

$$\mathbf{M} = \mathbf{r}_{cm} \wedge m\mathbf{g} + \mathbf{M}_{adj} + \mathbf{r}_{adj} \wedge \mathbf{R}_{adj}$$

where

$\mathbf{M}$  is the moment at the selected joint

$\mathbf{M}_{adj}$  is the moment at the adjacent joint

$\mathbf{r}_{adj}$  is the vector running from the selected joint to the adjacent one

$m$  is the mass of the segment between these two joints

$\mathbf{r}_{cm}$  is the vector running from the selected joint to the centre of mass of the segment

$\mathbf{R}_{adj}$  is the resultant force calculated at the adjacent joint

For example,

$\mathbf{R}_{adj}$  at the wrist is  $\mathbf{F} + m_{hand}\mathbf{g}$

$\mathbf{R}_{adj}$  at the elbow is  $\mathbf{R}_{wrist} + m_{forearm}\mathbf{g}$

and so on, if  $\mathbf{F}$  is a force on the hand.

## MOMENTS IN DYNAMIC TASKS

While previous sections in this chapter focused on calculations for static tasks, this section focuses on body segments that undergo accelerations and experience inertial forces that must be added to the moments at the joints. The main difficulty in performing an analysis of a dynamic task is recording instantaneous accelerations throughout the task. Early work by Ayoub and El Bassoussi (1978) included a prediction of accelerations in a dynamic computer model approximated as a function of the angle of each limb at the end of the lift and the duration of the lift. Motion tracking systems now offer accurate measures of the position, velocity and acceleration of targets.

The most common simplification is to ignore accelerations and treat the problem as a static one. This will lead to errors for many tasks. McGill and Norman (1985) carried out static and dynamic evaluations of L4–L5 intervertebral joint loading. Results with the dynamic analyses were on average 19% higher than with the static approximation, and some up to 52% higher. Garg et al. (1982) found that dynamic evaluations were two to three times higher than static ones. These differences are probably due to differences in lifting speed, method and weight lifted. McGill and



Norman (1985) proposed a quasi-dynamic model, in which the only dynamic component was the acceleration of the object lifted. Tsuang et al. (1992) have shown that differences between static, quasi-dynamic and dynamic analyses increase mainly with the speed of the lift. Danz and Ayoub (1992) found that peak vertical hand forces at the initiation of a fast lift were 3.0–3.5 times the magnitude of the load being lifted.

The moment at a particular joint varies throughout the motion due to changes both in lever arms and in accelerations. At any instant, the moment depends on the value of the linear acceleration of the centre of mass of each segment on the direction of this acceleration, and also on the angular acceleration of each segment. The resistance to rotation that an object has depends on its mass and shape and is described by its moment of inertia.

The general equation for the moment  $\mathbf{M}$  at a joint at a particular instant is

$$\mathbf{M} = \mathbf{r}_{\text{cm}} \wedge m\mathbf{g} + \mathbf{M}_{\text{adj}} + \mathbf{r}_{\text{adj}} \wedge \mathbf{R}_{\text{adj}} + \mathbf{r}_{\text{cm}} \wedge m\mathbf{a} + \mathbf{I} \wedge \ddot{\theta}$$

The first three terms have already been described (see Figure 17.6) in the previous section on static moments. Symbols used in the additional terms are the following:  $\mathbf{a}$  is the linear acceleration of the centre of mass of the segment ( $\text{m/s}^2$ ),  $\ddot{\theta}$  is the angular acceleration of the segment about its centre of mass ( $\text{degrees/s}^2$ ) and  $\mathbf{I}$  is the moment of inertia of the segment about its centre of mass ( $\text{kg m}^2$ ).

## PREDICTION OF STRENGTH AND TASK FEASIBILITY USING MOMENTS

Many experimental studies have been carried out to measure the maximum voluntary static strength (overall body strength and the strength of individual joints) of men and women. A prediction of the percentage of the population capable of performing a task can be useful for preplacement strength testing, return-to-work assessments and especially for job and product designers when simulating high-exertion tasks during the early part of the design process to avoid costly redesigns (see Chaffin, 1999).

One approach used for task assessment is to evaluate the moments required at each joint by a task and compare these moments with joint strength data from the population. At least 95% of the population should be accommodated by a given task design. Chaffin (1988) gives examples of this method; one example applies to pulling carts for moving stock, and it is shown how the percentage of women capable of this drops as the force required increases. Another example shows that most of the working population is capable of performing a particular lifting task – even though evaluation of loads on the spine shows the task is hazardous. A note of caution is warranted. Since these strength data were collected during static postures, they overestimate strength capacities during dynamic activities. Even though these are valuable design tools, the interpretation of results should consider the nature of the task.

## RELIABILITY OF MAXIMUM JOINT STRENGTH DATA

A problem with using joint strength data is the reliability of the data on a particular joint's maximum strength. A wide range of results can be found in the literature, and this may be due to different testing methods as well as to human variability. For instance, elbow strength results are sometimes given as the force participants are able to exert using their hands. This method is not satisfactory as the wrist may be the weak link in this task. To avoid this problem, there have been experiments in which the participants exert a force on a device attached proximal to the wrist. These results are incomplete if the distance between this point and the elbow is not known. The most useful data are those where elbow strength is expressed directly as a moment, and the method for its determination is noted. Low-back strength is particularly difficult to define because there is no obvious point from which to define moments when testing for strength.

### DEPENDENCE OF JOINT STRENGTH WITH JOINT ANGLES

A challenge for assessing joint strength relative to demands concerns joint angles. The force a muscle can exert depends on its length, so moment capabilities depend on joint angles. For example, elbow strength depends not only on the elbow angle but also on the shoulder angle as muscles span across both joints. Yet many results on elbow moments do not report the shoulder angle.

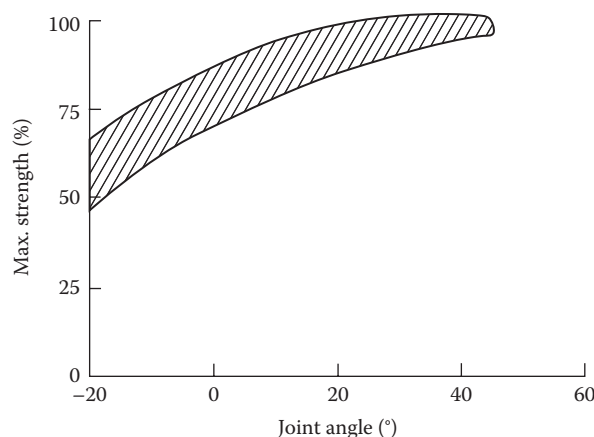
For a compilation from the literature of moment–angle curves of major joints, the reader is referred to Svensson (1987). Figure 17.7 has been adapted from one of these results and shows the dependence of the trunk extensor moment and trunk angle. The hatched area includes curves from four different studies.

A word of caution is in order when referring to published results concerning joint angles – the field of biomechanics does not appear to have any angle conventions and the position of the zero angle varies across studies. One useful standard may be that set by the British Orthopaedic Association (1966) in their booklet describing terminologies used in joint motion. Their method is the zero starting position, that is, to accept the anatomical position of a limb as zero degrees. For instance, the elbow angle is zero for the extended straight arm, and its range of movement is from about 150° flexion to 10° hyperextension (see Figure 17.8).

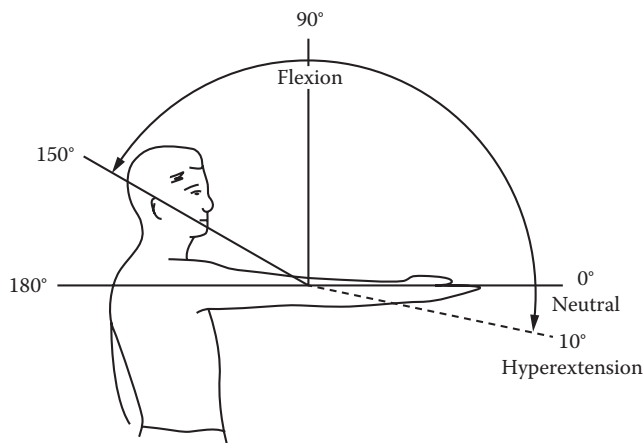
### INTERPRETATION OF RESULTS

Readers can refer to prediction equations in Chaffin et al. (1999, p. 263) for maximum moments as a function of joint angles. However, as a wide range of values can be found from other sources, a rough compilation of these ranges is presented in Table 17.1. This is only intended to give the reader an order of magnitude with which to compare evaluated moments, with further work needed in this area. The moment ranges given in Table 17.1 include all those found in some of the literature, and they all refer to the so-called fit and healthy volunteer. The angle notation in Table 17.1 follows the British Orthopaedic Association convention described earlier. Results refer to moments along a single axis (e.g. pure flexion or pure abduction), so caution must be exercised in using them for a task in which moments in several directions are combined.

An additional word of caution is warranted as antagonistic muscles (i.e. muscles that have opposite effects) often contract simultaneously to the agonist(s). For example, at some elbow angles, voluntary elbow flexion recruits not only the biceps (flexor) but also to some extent the triceps (extensor).



**FIGURE 17.7** Back extensor strength as a function of trunk angle. Results from four studies, normalised by denoting the top value of each curve as 100%. (From Svensson, O.K., On quantification of muscular load during standing work: A biomechanical study, Dissertation from the Kinesiology Research Group, Department of Anatomy, Karolinska Institute, Stockholm, Sweden, 1987.)



**FIGURE 17.8** Elbow – flexion and hyperextension. (From British Orthopaedic Association, *Joint Motion: Method of Measuring and Recording*, Published by the American Academy of Orthopedic Surgeons, 1966.)

**TABLE 17.1**  
**Maximum Voluntary Joint Strengths (N m) from Some of the Literature**

| Joint Strength       | Joint Angle<br>(Degrees) | Range of Moments<br>(N m) of Subjects<br>from Several Studies |        | Variation with Joint Angle                                                                      |
|----------------------|--------------------------|---------------------------------------------------------------|--------|-------------------------------------------------------------------------------------------------|
|                      |                          | Men                                                           | Women  |                                                                                                 |
| Elbow flexor         | 90                       | 50–120                                                        | 15–85  | Peak at about 90°                                                                               |
| Elbow extensor       | 90                       | 25–100                                                        | 15–60  | Peak between 50° and 100°                                                                       |
| Shoulder flexor      | 90                       | 60–100                                                        | 25–65  | Weaker at flexed angles                                                                         |
| Shoulder extensor    | 90                       | 40–150                                                        | 10–60  | Decreases rapidly at angles less than 30°                                                       |
| Shoulder adductor    | 60                       | 104                                                           | 47     | As angle decreases, strength increases and then levels at 30° to –301°                          |
| Trunk flexor         | 0                        | 145–515                                                       | 85–320 | Patterns differ among authors                                                                   |
| Trunk extensor       | 0                        | 143                                                           | 78     | Increases with trunk flexion                                                                    |
| Trunk lateral flexor | 0                        | 150–290                                                       | 80–170 | Decreases with joint flexion                                                                    |
| Hip extensor         | 0                        | 110–505                                                       | 60–130 | Increases with joint flexion                                                                    |
| Hip abductor         | 0                        | 65–230                                                        | 40–170 | Increases as angle decreases                                                                    |
| Knee flexor          | 90                       | 50–130                                                        | 35–115 | In general, decreases with knee flexion but some disagreement with this, depending on hip angle |
| Knee extensor        | 90                       | 100–260                                                       | 70–150 | Minima at full flexion and extension                                                            |
| Ankle plantarflexor  | 0                        | 75–230                                                        | 35–130 | Increases with dorsiflexion                                                                     |
| Ankle dorsiflexor    | 0                        | 35–70                                                         | 25–45  | Decreases from maximum plantar flexion to maximum dorsiflexion                                  |

*Note:* The ranges presented include the ranges from these studies.

This co-contraction is often believed to have a joint stabilising role as one effect is increased joint stiffness. Therefore, it is possible that prediction models overestimate a participant’s strength if co-contraction is beyond the level during strength testing.

Currently, there are very few data on maximum moments exerted dynamically, probably because of the problem in recording moments that cover a wide enough maximum dynamic range of angles, velocities and accelerations. In conclusion, comparing the moments required by a task with data

on maximum capabilities provides some useful guidance in terms of orders of magnitude, but the method should be used with caution for the reasons noted here.

## FORCES ON THE LOW BACK

A study by Chaffin (1988) was mentioned earlier, in which a particular lifting task was analysed. From the moments at the joints, it was estimated that most of the working population would be capable of performing the task; however, evaluation of loads on the spine indicated that the task could put the back at risk. This section discusses criteria available for assessing such risk. A simple 2D model to calculate the compression force on the spine is shown in Figure 17.3. This section will describe other models, including 3D ones for asymmetrical postures.

## GUIDELINES FROM LOW-BACK FORCES

The most commonly used guideline for task assessment is the value of the compression force between vertebrae. Experiments on cadaveric spines have shown that fractures appear above certain levels of compression. The level is lowest for older people; female spines are, as a general rule, weaker than male spines (Hutton and Adams, 1982). The National Institute for Occupational Safety and Health (NIOSH, 1981) concluded that tasks causing a compression on the lumbosacral joint greater than 6400 N are above the NIOSH's (1981) 'maximum permissible limit', making the tasks unacceptable and requiring engineering controls reduce compression forces. On the other hand, compressions under 3400 N can be tolerated by most young, healthy workers (over 75% of women and over 99% of men). The revised NIOSH equation uses a single 3400 N criterion (Waters et al., 1993). It must be noted that these guidelines relate to lifting in the sagittal plane, and the spine may be much more vulnerable under axial rotation or hyperflexion (Adams and Hutton, 1981). The majority of compression tests have sought the ultimate compression strength, but work by Brinckmann et al. (1987) provides data relating to repetitive tasks and the strength of intervertebral joints under cyclic loading. For instance, they have shown that for a cyclic load of about half the ultimate compression strength, the probability of a fatigue fracture after 100 cycles is nearly 50%. The value of the compression force may not be the most relevant parameter related to back injury, and guidelines should be used with their limitations in mind, especially at extremes of trunk motion and in dynamic tasks.

Jäger and Luttmann (1999) argue that the 3400 N criterion does not have sufficient epidemiological or biomechanical justification. Jäger and Luttmann (1997) recommend compression limits between 1.8 and 6.0 kN, depending upon the age and gender of the worker. There is not widespread consensus regarding the use of a specific criterion, and questions have been raised regarding the applicability of cadaveric data to set *in vivo* limits (Dempsey, 1998). Factors such as the temperature of the specimens, thawing effects, specimen fixation and the testing environment (Adams, 1995) affect the strength of spinal segments. However, spinal compression estimates provide insight into the task demands at the low back and provide a basis for comparison of different design alternatives.

## Applications of Low-Back Modelling

The value of lumbar spine compression is the most frequently used criterion in the evaluation of tasks that may put the back at risk. This criterion has been used extensively in the analysis of lifting tasks, for example, to compare lifting techniques (Chaffin, 1999), evaluate patient-handling techniques (Gagnon et al., 1986), compare welding systems (Fethke et al., 2011) or to determine maximum acceptable weights (Hutton and Adams, 1982; Jäger and Luttmann, 1986). More examples of the use of low-back modelling in the analysis of industrial tasks can be found in Norman and McGill (1999).

Due to the inherent difficulty in collecting postural data in the workplace, one approach is to use simulation to generate 2D motion data from four frames of a standard videotape of a lift (Chang et al., 2003). The postures in the four frames are matched with a mannequin in a computer program

to generate the initial input data. The use of four frames was suggested by Hsiang et al. (1998) as an optimum number (fewer frames degraded accuracy of the predictions, whereas more frames did not significantly enhance accuracy). A motion pattern prediction algorithm (Hsiang et al., 1999) is then used to simulate reasonably accurate motion data for the entire lift, and temporal information is retrieved from the videotape (based on frame rate of the input video). The advantage of this approach is that the analysis approximates closely the results provided by a dynamic analysis conducted in the laboratory.

Xu et al. (2012) extended the approach used by Chang et al. (2003) to cover asymmetric lifts using four-point cubic spline interpolation of segment Euler angles combined with a biomechanical model. The authors concluded that the results provided acceptable estimates of spinal compression without expensive data acquisition.

Ergonomists often assess peak loads in individual tasks. Peak loads may be appropriate for some analyses, but cumulative loads may be more relevant, for instance, with respect to low-back pain (Norman et al., 1998). Assessment of cumulative loads requires the continuous recording of information on 'all' tasks in the job, including those with no manual materials handling. Few investigators have assessed cumulative loads (Kumar, 1990; Norman et al., 1998; Jäger et al., 2000), likely because the data collection and reduction costs are not trivial.

## MODELS OF THE LOW BACK

There were several early studies to evaluate forces on the spine through in vivo measurement (e.g. Nachemson and Morris, 1964; Nachemson and Elfström, 1970). Following these studies, intradiscal pressure measurements have been used mostly in conjunction with electromyography (EMG) and intra-abdominal pressure (IAP) measurements to verify the validity of low-back models (Schultz et al., 1982). Such direct measurement is rarely practical or allowed by ethics committees, leading to modelling efforts discussed later.

In the simple example of Figure 17.3, one set of back muscles, situated posterior to the spine, is used to resist a trunk flexion moment. This created a compression force on the spine. This is only one model of the low back; however, many others, with varying complexity, have been developed. The rest of this section describes how to evaluate muscle and compression forces with 2D and 3D models; these sections are provided for readers who wish to carry out calculations themselves.

### A Simple 2D Model

We will now complete the simple 2D model (in Figure 17.3) used at the beginning of this chapter. This model allows back muscles (representing the erector spinae group) to resist a trunk flexion moment. If trunk extension is to be resisted, these muscles can relax and let abdominal muscles (rectus abdominis) take over. Figure 17.9 summarises the results of this model. The lumbar spine compression is shown as a function of the low-back moment. The vertical force on the hands also compress the spine, as shown from the parallel lines for 0, 1000 N upwards and 1000 N downwards.

The equations describing this 2D model, and from which Figure 17.9 was obtained, are described later. A flexion moment is provided by the rectus abdominis ( $R$ ), and an extension moment is provided by the erector spinae ( $E$ ) (see Figure 17.10).  $E$  acts at a distance  $y_E$  from the centre of the spine, and the lever arm for  $R$  is  $y_R$ .

If  $M_x$  is the flexion or extension moment which these muscles must provide and,

if  $M_x > 0$  (flexion required), then

$$M_x = (-y_R) \times (-R)$$

(both the  $y$  and  $z$  axes are in an opposite direction to  $y_R$  and  $R$ , hence the minus signs)

$$\text{so } M_x = y_R R$$

else if  $M_x < 0$  (extension required)

$$M_x = -y_E E$$

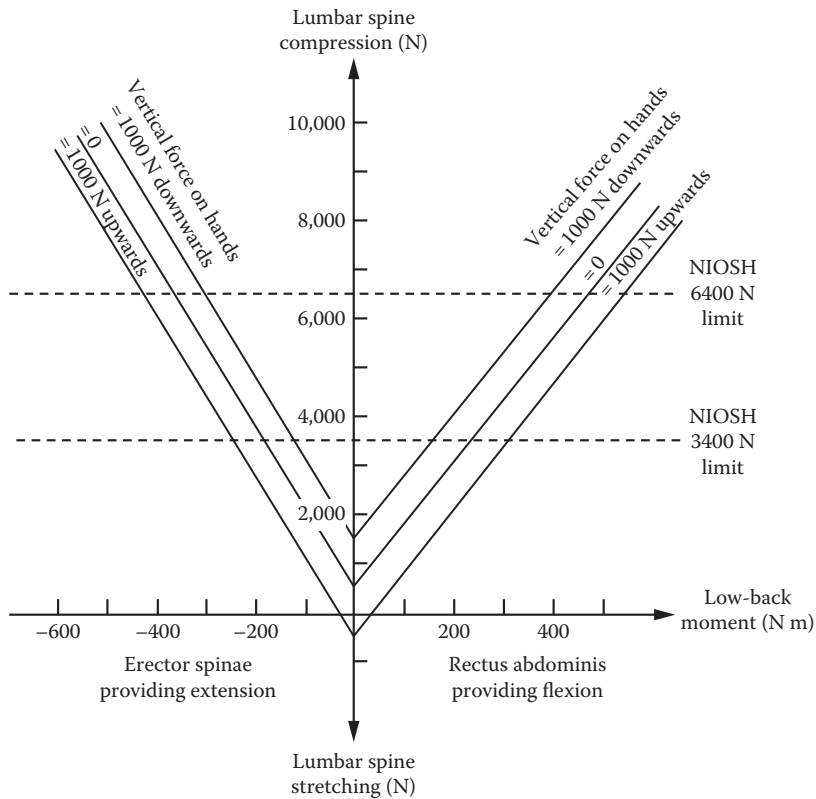


FIGURE 17.9 Lumbar spine compression as a function of low-back moment, with a simple 2D model.

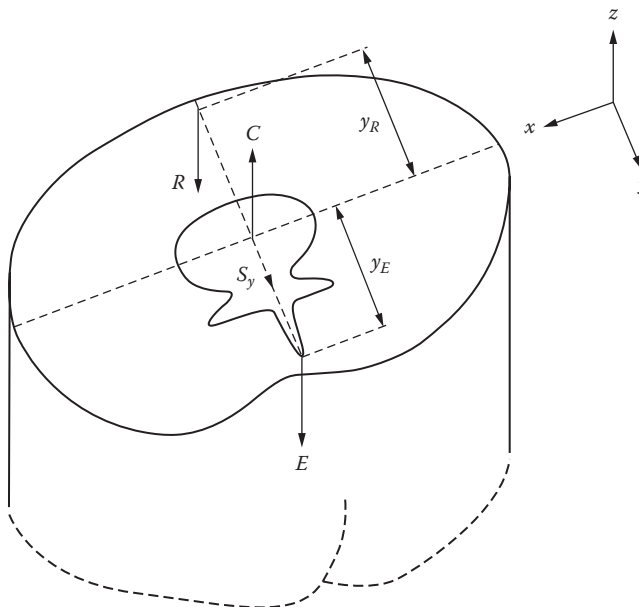


FIGURE 17.10 A simple 2D model represented on a section of the low back ( $R$ , rectus abdominis;  $E$ , erector spinae;  $C$ , compression;  $y_R$ ,  $y_E$ , lever arms).



Finally,  $E$ ,  $R$  and  $C$  must add up to a vertical force  $F_z$  that counteracts the weight of the body and any external downwards force acting at the hands:

$$F_z = -(\text{sum of body weights and vertical hand force})$$

$$F_z = C - R - E$$

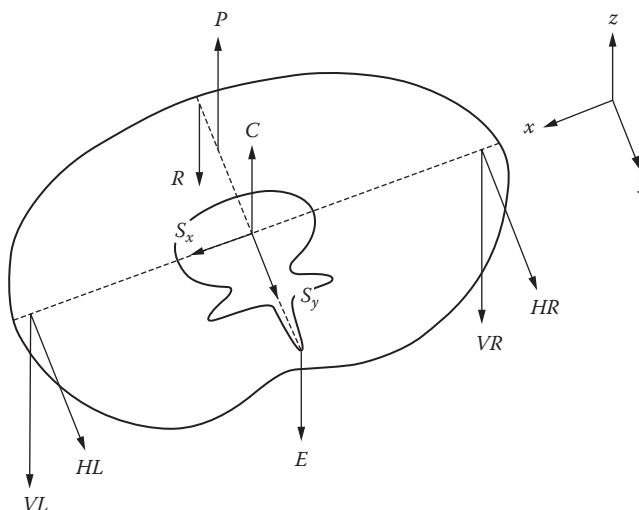
so the value of  $C$ , the compression, is obtained.

In the same way, any horizontal force at the hands will be compensated by a horizontal shear force  $S_y$  at the intervertebral joint.

For flexion/extension moments, this crude model gives estimates that are very similar to those given by a more detailed model. Returning to the graphical summary in Figure 17.9 for flexion/extension, the slopes on the graph are  $1/y_R$  for positive moments and  $1/y_E$  for negative moments, with  $y_R = 8.0$  cm and  $y_E = 5.8$  cm. (These numbers are average values; see section 'Low-Back Geometry'.)

### Three-Dimensional Models

The previous model did not have any muscles accounting for lateral flexion or axial rotation of the trunk. A model described by Chaffin et al. (1999, p. 239) will be briefly summarised. The model consists of six muscles (Figure 17.11) including the rectus abdominis ( $R$ ) and erector spinae ( $E$ ), which provide flexion and extension, respectively; the vertical components of the left and right obliques ( $VL$  and  $VR$ ), which provide lateral flexion to the left and to the right; and the horizontal components of the left and right obliques ( $HL$  and  $HR$ ), which provide positive (anticlockwise) and negative axial moments around the  $z$  axis. Other low-back forces in the model are the compression force  $C$  on the intervertebral joint (if  $C < 0$ , the force is on the contrary an extension force), and the lateral and anteroposterior shear forces,  $S_x$  and  $S_y$ , on the intervertebral joint. One more force not mentioned so far is the force  $P$  due to IAP; it is believed that the rise in pressure in the abdominal cavity that occurs during heavy manual handling supports the trunk and effectively produces an extensor moment. This moment is equivalent to a force  $P$  acting on the centre of the diaphragm. This topic will be discussed in more detail later.



**FIGURE 17.11** Schematic diagram of a simple 3D model. ( $R$ ,  $E$ ,  $C$ , rectus abdominis, erector spinae and compression;  $P$ , force due to intra-abdominal pressure;  $VL$ ,  $HL$  are the vertical and horizontal components of the obliques on the left side of the body;  $VR$ ,  $HR$  on the right side.) (Adapted from Chaffin and Andersson, 1984.)

**TABLE 17.2**  
**Equations for the Simple 3D Model**

|          |        |                           |           |     |   |                                    |
|----------|--------|---------------------------|-----------|-----|---|------------------------------------|
| IF $M_x$ | $\geq$ | 0                         | THEN $E$  | $=$ | 0 | (Flexion required)                 |
| $M_x$    | $\leq$ | 0                         | THEN $R$  | $=$ | 0 | (Extension required)               |
| IF $M_y$ | $\geq$ | 0                         | THEN $VR$ | $=$ | 0 | (Flexion to left required)         |
| $M_y$    | $\leq$ | 0                         | THEN $VL$ | $=$ | 0 | (Flexion to right required)        |
| IF $M_z$ | $\geq$ | 0                         | THEN $HR$ | $=$ | 0 | (Anti clockwise rotation required) |
| $M_z$    | $\leq$ | 0                         | THEN $HL$ | $=$ | 0 | (Clockwise rotation required)      |
| $F_x$    | $=$    | $S_x$                     |           |     |   |                                    |
| $F_y$    | $=$    | $S_y + HL + HR$           |           |     |   |                                    |
| $F_z$    | $=$    | $C + P - R - E - VL - VR$ |           |     |   |                                    |
| $M_x$    | $=$    | $-y_R P + y_R R - y_E E$  |           |     |   |                                    |
| $M_y$    | $=$    | $x_p (VL - VR)$           |           |     |   |                                    |
| $M_z$    | $=$    | $-x_p (HL - HR)$          |           |     |   |                                    |

Source: Adapted from Chaffin and Andersson (1984).

The equations for this model are presented in Table 17.2. Anteroposterior lever arms are denoted by  $y$ , lateral lever arms by  $x$ , so  $y_E$  is the distance of the erector spinae force  $E$  behind the spine, for example  $x_O$  is the lateral lever arm of the obliques.  $M_x$ ,  $M_y$  and  $M_z$  are the low-back moments to be provided by the muscles.  $F_x$ ,  $F_y$  and  $F_z$  are the forces provided by the low back to counteract body weight and hand forces.

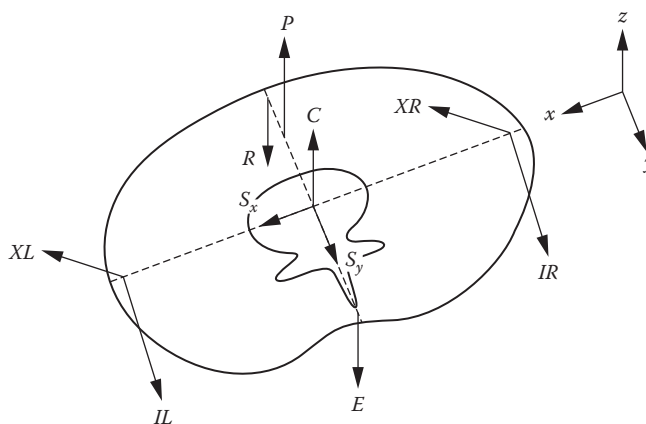
One limitation of this model is that the vertical and horizontal components of the obliques are made to act independently, whereas oblique muscles in reality always pull simultaneously in the vertical and horizontal directions. As this model allows an oblique to provide a purely horizontal force, the compression on the spine may be underestimated for tasks involving axial rotation. We will discuss later a 10-muscle model by Schultz and Andersson (1981), which models the obliques in a more realistic way, with internal obliques acting posteriorly downwards and external obliques acting anteriorly downwards.

This 10-muscle model, and others involving more muscles, requires a computer to carry out a particular mathematical procedure (linear optimisation), indicating that such calculations are not trivial. Accordingly, a micro-model was proposed by Tracy (1988) to model the obliques more realistically and produce results that are closer to those of models requiring linear optimisation.

The rule for oblique action is as follows:

Internal and external obliques pull both posteriorly and anteriorly downwards, as in Figure 17.12. Suppose a clockwise axial rotation moment must be provided ( $M_z < 0$ ); this can be done by the external on the left ( $XL$ ) and by the internal on the right ( $IR$ ). If lateral flexion to the left is also required ( $M_y > 0$ ),  $XL$  and  $IL$  will be in action. With this model,  $XL$  acts strongly to provide both  $M_z$  and  $M_y$ , and either  $IR$  or  $IL$  act, depending on which moment is the highest. So, if axial rotation is more important than lateral flexion,  $XL$  and  $IR$  are active.

The equations for this model are given in Table 17.3, which contains predictions that come close to those of the 10-muscle model that is to follow and for which a computer is not needed. The main limitations of these equations are that the erector spinae and the rectus abdominis ( $E$  and  $R$ ) are placed in the mid-sagittal spine, whereas in reality they are groups of muscles situated to the left and the right. They contribute to lateral flexion moments, whereas the simple models in Table 17.3 only allow obliques to contribute.



**FIGURE 17.12** Schematic diagram of the micro-model – a simple 3D model with more realistic representation of the obliques (symbols described in Table 17.3).

### 3D Models Requiring Optimisation

If the erector spinae and the rectus abdominis are placed as separate forces on either side of the sagittal plane, and if any other trunk muscles are also represented, special techniques are required to estimate forces exerted by the different muscles. Lateral flexion to the right, for instance, can now be provided by the right obliques, the right erector spinae or the right rectus abdominis. Mathematically, there are more variables (forces) than equations. The indeterminacy must be solved by making assumptions; giving rules as in Table 17.2 or 17.3 is not possible or practical when many muscles are involved.

An alternate technique is then to use linear programming that optimises one variable while making sure a number of equations are satisfied. Schultz and Andersson (1981) and Schultz et al. (1983) have established models with 10–22 muscles based on the assumption that the spinal compression  $C$  is at a minimum. A linear programming routine will ensure that all muscle forces provide the required moments and do not exceed a maximum capability of 100 N/cm<sup>2</sup> and that, at the same time, the forces have been distributed so that  $C$  is as low as possible. Although the problem may seem complicated, linear programming routines are in principle straightforward to use. For asymmetrical tasks, models such as the ones developed by Schultz et al. (1983) are far superior to simple calculator-based models, as shown by Schultz et al. in validation experiments.

The condition to minimise the compression force will have the effect that a muscle with a larger lever arm will provide a force in preference to one with a smaller lever arm. Muscles close to the spine act only if other muscles have reached their maximum capability. Bean et al. (1988) proposed a double linear programming alternative that includes first using linear programming to minimise maximum muscle intensities (the optimal intensity value from the solution was denoted  $I^*$ ). This ensures that no muscle is giving its maximum while other muscles, which could also contribute, are inactive. The next step is to solve the problem all over again, minimising  $C$ , but with the condition that no muscle intensity exceeds  $I^*$ .

Hughes and Chaffin (1995) compared the Bean et al. (1988) approach with that of minimising the sum of the cubed muscle stresses using EMG data collected from eight participants while resisting extension and torsion moments. The authors did not find one approach to be superior for all muscles as the sum of the cubed stress model predicted rectus abdominis activity the best, whereas the Bean et al. approach predicted left erector spinae, right and left external oblique, and left latissimus dorsi activity better.

**TABLE 17.3**  
**Equations for the Micro-Model Shown in Figure 17.12**

*Connections*

|                                                    |                                                                                                                                                   |
|----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>R</i>                                           | Rectus abdominis                                                                                                                                  |
| <i>E</i>                                           | Erector spinae                                                                                                                                    |
| <i>XL, XR</i>                                      | Left and right external obliques, acting in the (y, z) plane, downwards and towards the ventral part of the trunk, at 45° to the transverse plane |
| <i>IL, IR</i>                                      | Left and right internal obliques, acting in the (y, z) plane, downwards and towards the dorsal part of the trunk, at 45° to the transverse plane  |
| <i>P</i>                                           | Force due to intra-abdominal pressure                                                                                                             |
| <i>C, S<sub>x</sub>, S<sub>y</sub></i>             | Compression and shear forces on the intervertebral joint                                                                                          |
| Axes (x, y, z)                                     | Centred on the spine. (x, y) in the transverse plane with x: to the left of the body, y directed posteriorly, z directed upwards                  |
| <i>F<sub>x</sub>, F<sub>y</sub>, F<sub>z</sub></i> | Resultant reaction forces at the level of the section                                                                                             |
| <i>M<sub>x</sub>, M<sub>y</sub>, M<sub>z</sub></i> | Resultant reaction moments at the level of the section                                                                                            |
| ABS ( <i>x</i> )                                   | Absolute (positive) value of <i>x</i>                                                                                                             |
| SUM =                                              | $\frac{M_y + M_x}{2(x_0 \cos 45)}$                                                                                                                |
| DIFF                                               | $\frac{M_y - M_x}{2(x_0 \cos 45)}$                                                                                                                |

*Equations*

$$F_x = C + P - E - R - (IL + IR) \cos 45 - (XL + XR) \cos 45$$

$$F_y = (IL + IR) \sin 45 - (XL + XR) \sin 45 + S_y$$

$$F_x = S_z$$

$$M_x = x_0(IL - IR) \cos 45 + x_0(XR + XL) \sin 45$$

$$M_y = x_0(IL - IR) \sin 45 + x_0(XR + XL) \sin 45$$

$$\text{If } M_x \geq 0 \text{ then } E = 0$$

$$\text{If } M_x < 0 \text{ then } R = 0$$

$$\text{If } M_y > 0 \text{ and } M_z > 0$$

$$\text{and } \text{ABS}(M_y) \geq \text{ABS}(M_z) \text{ then}$$

$$IR = 0$$

$$XR = 0$$

$$IL = \text{SUM}$$

$$XL = \text{DIFF}$$

$$IR = 0$$

$$XR = 0$$

$$IL = -\text{SUM}$$

$$XL = -\text{DIFF}$$

$$\text{and } \text{ABS}(M_y) \geq \text{ABS}(M_z) \text{ then}$$

$$\text{If } M_y \geq 0 \text{ and } M_z \leq 0$$

$$\text{and } \text{ABS}(M_y) \geq \text{ABS}(M_z) \text{ then}$$

$$IR = 0$$

$$XR = 0$$

$$IL = \text{SUM}$$

$$XL = \text{DIFF}$$

$$\text{and } \text{ABS}(M_y) \geq \text{ABS}(M_z) \text{ then}$$

$$IR = 0$$

$$XR = 0$$

$$IL = -\text{SUM}$$

$$XL = -\text{DIFF}$$

(Continued)

**TABLE 17.3 (Continued)****Equations for the Micro-Model Shown in Figure 17.12**If  $M_y < 0$  and  $M_z < 0$ and  $\text{ABS}(M_y) \geq \text{ABS}(M_z)$  then

$$IL = 0$$

$$XL = 0$$

$$IR = -\text{SUM}$$

$$XR = -\text{DIFF}$$

and  $\text{ABS}(M_y) < \text{ABS}(M_z)$  then

$$IL = 0$$

$$XR = 0$$

$$IR = -\text{SUM}$$

$$XL = \text{DIFF}$$

If  $M_y < 0$  and  $M_z \geq 0$ and  $\text{ABS}(M_y) \geq \text{ABS}(M_z)$  then

$$IL = 0$$

$$XL = 0$$

$$IR = -\text{SUM}$$

$$XR = -\text{DIFF}$$

and  $\text{ABS}(M_y) < \text{ABS}(M_z)$  then

$$IR = 0$$

$$XL = 0$$

$$IL = -\text{SUM}$$

$$XR = -\text{DIFF}$$

In a related study, Nussbaum et al. (1995) compared the ability of an artificial neural network (ANN) to predict muscle activity as compared to the approaches discussed in the previous paragraph. The ANN results were more highly correlated with the EMG activity, indicating that this approach provides an alternative to either optimisation or actual EMG input.

### Three-Dimensional Models Requiring Biological Input

Some biomechanical models require the input of EMG signals, known as biological input. This approach bypasses the optimisation and ANN predictions discussed earlier and will only be mentioned briefly. EMG-assisted models are often very complex and require exceptional expertise not only due to the biomechanical modelling but also to properly acquire, treat and interpret EMG data from participants (or workers) performing real tasks. As such, the models are rarely used by non-researchers. An overview of the model developed at the University of Waterloo, and additional references, is provided by McGill (1999). Details of the model developed at Ohio State, and additional references, can be found in Marras and Granata (1997).

There are tradeoffs when choosing between EMG data and optimisation approaches. Although optimisation requires less data collection and equipment, EMG may be able to provide a more realistic view of activation patterns, especially given the high between-subject differences in trunk muscle co-activation that have been observed (e.g. Perez and Nussbaum, 2003). EMG and optimisation approaches can also be combined into an EMG-assisted optimisation approach (Cholewicki et al., 1995).

## INPUTS TO BIOMECHANICAL CALCULATIONS

### POSTURE INPUT

Biomechanics is sometimes used as a predictive tool on a posture that has not been observed but has been estimated as a likely posture for a task. However, the posture may not be realistic and it may be worth ensuring that the body is in balance by checking that the resultant of all external

forces lies in the area between the 2 ft. If the posture is asymmetric, each leg can take a different proportion of the resultant force at the feet, and this needs to be measured with a force plate, which brings us back to the laboratory. Therefore, biomechanics must be used with caution when used as a predictive tool.

### **BODY SEGMENT WEIGHTS**

Table 17.4 summarises masses and the locations of the centre of gravity of body segments compiled by Pheasant (1986). Other anthropometric data used in modelling, such as link lengths, can be found in the same reference. Segment mass data originate from very small, poor representative samples, so may be a source of error in biomechanical calculations.

### **LOW-BACK GEOMETRY**

There have been recent improvements in the data available for low-back models. Both computed axial tomography (CAT) and magnetic resonance imaging (MRI) have been used to measure muscle lever arms and cross-sectional areas, whereas previously, values came from a limited sample of cadaveric data. Some of the results needed for the models described in this chapter can be found in Table 17.5. These results are lever arms for 96 females (Chaffin et al., 1990) and at the L3–L4 level for 26 males (Tracy et al., 1989). Data for more muscles and at other lumbar sections can be found in these same studies. Lever arms for both sexes have also been measured by Nemeth and Ohlsen (1986) at the lumbosacral joint and by Kumar (1988) at L3 and L4 (as well as T7 and T12). Daggfeldt and Thorstensson (2003) reported MRI results for lever arms between T12/L1 and L5/S1 for four participants in two different postures.

### **ROLE OF INTRA-ABDOMINAL PRESSURE**

The most widespread theory of the role of IAP in low-back force production is that the pressure supports the trunk, and IAP's action on the diaphragm and pelvic floor is equivalent to a force for trunk extension. According to this model, the force produced by IAP is calculated by multiplying the pressure by the area of the diaphragm. The extensor moment created by this force is the product of the force with the lever arm of the centroid of the area on which IAP acts. Using this model, IAP reduces lumbar compression by 4%–30% according to Schultz et al. (1982) and by 2%–8% according to Leskinen and Troup (1984). Recently, Daggfeldt and Thorstensson (2003) reported that IAP produced 9%–13% of the extension torque during static exertions.

The reduction in spinal compression due to IAP may be underestimated because calculations of moments and lumbar loads, ignoring IAP, still sometimes result in excessive compression values although no structural failure is observed (Jones, 1983). However, both Krag et al. (1985) and Nachemson et al. (1986) have argued from experimental evidence that IAP does not reduce lumbar compression; EMG readings showed that trunk extensor muscle action was not reduced when the abdominal cavity was voluntarily pressurised.

A number of theories for the role of IAP have been put forward, and the reader is referred to Aspdén (1987) for a review of these various theories. McGill (1999) indicated that IAP may not have a direct role in reducing spinal compression, but rather may act to increase trunk stiffness to prevent tissue strain. Until the controversies on the role of IAP are resolved, one approach is to represent IAP as an extensor force, as in Figure 17.11. Shown in Table 17.6 are some values found in the literature for various tasks, but many authors choose to ignore IAP and set the force to zero. It has been possible to measure IAP with a swallowed radio pill for some time (Davis and Stubbs, 1977), but some experience is required to use this technique. Chaffin et al. (1999, p. 227) have published a prediction equation for IAP using the hip moment and angle for lifting in the sagittal plane.



TABLE 17.4  
Segment Masses and Locations of Centre of Gravity

| Segment                           | Mass (Percentage Body Mass) | Location of Centre of Gravity                                                                                                                              |      |      |      |
|-----------------------------------|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------|------|
| 1. Head and neck                  | 8.4                         | 57% of distance from C7 to vertex                                                                                                                          |      |      |      |
| 1a. Head                          | 6.2                         | 20 mm above tragon                                                                                                                                         |      |      |      |
| 2. Head and neck and trunk        | 58.4                        | 40% of distance from hip to vertex                                                                                                                         |      |      |      |
| 2a. Trunk                         | 50.0                        | 46% of distance from hip to C7                                                                                                                             |      |      |      |
| 2b. Trunk above lumbosacral joint | 36.6                        | 63% of distance from hip to C7                                                                                                                             |      |      |      |
| 2c. Trunk below lumbosacral joint | 13.4                        | Approximately at tile hip joint                                                                                                                            |      |      |      |
| 3. Upper arm                      | 2.8                         | 48% of distance from shoulder to elbow joints                                                                                                              |      |      |      |
| 4. Forearm                        | 1.7                         | 41% of distance from elbow to wrist joints                                                                                                                 |      |      |      |
| 5. Hand                           | 0.6                         | 40% of hand length from wrist joint (at centre of an object gripped)                                                                                       |      |      |      |
| 6. Thigh                          | 10.0                        | 41% of distance from hip to knee joints                                                                                                                    |      |      |      |
| 7. Lower leg                      | 4.3                         | 44% of distance from knee to ankle joints                                                                                                                  |      |      |      |
| 8. Foot                           | 1.4                         | 47% foot length forward from the heel (half height of ankle joint above the ground) – mid-way between ankle and toll of foot at the head of metatarsal (I) |      |      |      |
| Total body mass <sup>a</sup> (kg) | Men                         | Women                                                                                                                                                      |      |      |      |
| Percentiles                       | 5th                         | 5th                                                                                                                                                        | 5th  | 95th | S.D. |
| British (19–65 years)             | 55.3                        | 74.5                                                                                                                                                       | 93.7 | 62.5 | 11.2 |

Source: From Pheasant, S., *Bodyspace: Anthropometry, Ergonomics and Design*, Taylor & Francis Group, London, U.K., 1986.

<sup>a</sup> Masses (in kg) to be multiplied by 9.81 to obtain weights (or forces in N) for the calculation of moments.

Source: From Pheasant, S., *Bodyspace: Anthropometry, Ergonomics and Design*, Taylor & Francis Group, London, U.K., 1986.

<sup>a</sup> Masses (in kg) to be multiplied by 9.81 to obtain weights (or forces in N) for the calculation of moments.

**TABLE 17.5**

**Lever Arms of Some Muscles at L3–L4 Level, from a CT Study of 96 Women (Chaffin et al., 1990) and in MRT Student of 26 Males (Tracy et al., 1989)**

|                                | Erector Spinae | Rectus Abdominis | Oblique  |
|--------------------------------|----------------|------------------|----------|
| <i>Females</i>                 |                |                  |          |
| Anteroposterior lever arm (mm) | 52 (4)         | 70 (19)          | 20 (10)  |
| Lateral lever arm (mm)         | 34 (4)         | 43 (11)          | 113 (16) |
| <i>Males</i>                   |                |                  |          |
| Anteroposterior lever arm (mm) | 58 (5)         | 80 (18)          | 17 (12)  |
| Lateral lever arm (mm)         | 38 (3)         | 34 (10)          | 122 (11) |

*Note:* Standard deviations in parentheses.

**TABLE 17.6**

**Intra-Abdominal Pressure for Various Tasks (1 kPa = 7.6 mmHg)**

| Task                                            | IAP (kPa) | Force (N) Developed by IAP over 299 cm <sup>2</sup> |
|-------------------------------------------------|-----------|-----------------------------------------------------|
| Schultz et al. (1982)                           |           |                                                     |
| Relaxed standing                                | 1.0       | 30                                                  |
| Uprights, arms in, holding 8 kg in both hands   | 1.5       | 45                                                  |
| Flexed 30°, arms out                            | 4.2       | 125                                                 |
| Flexed 30°, arms out holding 8 kg in both hands | 4.4       | 130                                                 |
| Davis and Stubbs (1977)                         |           |                                                     |
| Breathing                                       | 1         | 30                                                  |
| 90 mmHg 'safe limit'                            | 12        | 560                                                 |
| Grieve and Pheasant (1982)                      |           |                                                     |
| Competitive weightlifting                       | 40        | 1196                                                |
| Nachemson et al. (1986)                         |           |                                                     |
| Valsalva manoeuvre                              | 4         | 120                                                 |

Estimate for diaphragm area: 299 cm (Leskinen and Troup, 1984).

Estimate for IAP lever arm: 48 mm (Schultz et al., 1982).

## ANGLE OF DISCS

So far we have referred to forces on the lumbar spine without specifying a particular vertebra or disc. The low-back models discussed are too crude to differentiate between different vertebrae, and the main difference between calculations at L3 or at L5–S1 is the weight of the trunk above it. There is one other difference, though, and that is the angle of the intervertebral discs. The force referred to as the compression force earlier is actually partly compression and partly shear if the intervertebral joint is not perpendicular to the line of action of the erector spinae or rectus abdominis. Unfortunately, there is very little information on disc angles for various postures, and as there are large variations in the degree of lordosis in the population, predictions on disc angles are associated with a large uncertainty. One approach is to infer the shape of the spine from the shape of the surface of the back (Stokes and Moreland, 1987; Tracy et al., 1989), but this work is still somewhat inconclusive. Until more information is available, one solution is to make the approximation that L3 remains perpendicular to the line of action of the erector spinae and rectus

abdominis, whatever the posture (e.g. Schultz et al., 1983). For other approaches, and examples of how these angles have been derived for modelling, see Chaffin et al. (1999, p. 225). In any case, the uncertainty will mean that some of the compression force evaluated may be, in fact, a shear force, and vice versa.

## USES AND LIMITATIONS OF BIOMECHANICS

This chapter has surveyed the application of biomechanics in the E/HF field. Calculating moments at joints provides an estimate of the demands of the task. Moments calculated in 3D will also highlight possible twisting efforts that could be eliminated. Biomechanical calculations allow applied forces or posture to be varied so that problems and solutions can be identified.

## REPETITIVE WORK AND FATIGUE

Biomechanics cannot on its own answer questions such as, What force can be applied safely and without fatigue  $x$  times a minute for  $y$  hours, given  $n$  rest pauses of  $m$  minutes are provided? Biomechanics is better suited at providing estimates of instantaneous forces. Physiology can be helpful when evaluating the fatigue potential of tasks (localised and whole body). Even less can be said currently about dynamic work, presumably because of the large number of variables in the problem. In general, whether the work is static, intermittent or dynamic, biomechanics cannot on its own give reliable answers, except in extreme cases where the task can be shown to be so strenuous it can only be performed occasionally.

## SOURCES OF INACCURACY

Results of low-back forces vary across models, so it is useful to bear in mind the assumptions and simplifications a model employs. Biomechanical models should not be used unless the user is well aware of the simplifying assumptions and how they impact interpretation of the results.

Some inputs to low-back calculations are subject to uncertainty. Data acquisition can be an initial source of error as no posture recording system is completely accurate. The problem of predicting or measuring muscle activity to deal with indeterminacy was discussed earlier as some error will be present regardless of the approach used. Currently, some estimates of the weights and centres of mass of body segments are often required, and these are necessarily inexact estimates based primarily on studies of cadavers.

The role of IAP remains unclear, and the lack of direct measurement will lead to estimation errors. However, the effect of this potential inaccuracy may be attenuated because comparative calculations are prone to the same error.

There are individual variations in spinal shape between participants, and there is further uncertainty on disc angle changes with trunk motion. The highest compression estimates will be obtained with discs that are perpendicular to the line of action of the muscles. There is high inter-participant variation in some muscle lever arms; for example, in Table 17.5, the lever arm of the erector spinae has a standard deviation of nearly 10% of the mean value. Accordingly, the uncertainty about the force provided by the erector spinae is also represented by a standard deviation of about 10% of the force calculated with the mean lever arm.

All these sources of uncertainty are not usually critical to the interpretation when results are used to compare tasks (unless the magnitude of the error differs due to the condition being investigated), but are useful to bear in mind when results are used as absolute numbers, and perhaps evaluated against guidelines. There is a mathematical method of evaluating the effect of all the uncertainties on the final result (Barford, 1985), but another way is to experiment with different values of the input parameters to ascertain the sensitivity of the model.

## CONCLUSIONS

Biomechanics is a useful tool to evaluate manual handling tasks, highlight problems and compare possible alternate task designs. The models that are used can be more or less sophisticated; some require computers, although much evaluation can be achieved just with a calculator. Oftentimes, practical constraints in the workplace limit the sophistication of biomechanical analyses. Although more sophisticated methods can have higher validity, practical considerations can limit their applicability to some circumstances.

Results should be interpreted with a basic knowledge of the simplifications and uncertainties that were involved in the calculations. Even if one uses a software package available commercially, an understanding of the limitations of the model and required assumptions is necessary for a competent application. Other methods that are used to complement biomechanics include physiological measurements, EMG, injury statistics, discomfort charts and questionnaires.

## ACKNOWLEDGEMENT

The authors are indebted to Diva Ferreira for her contribution to the development of this chapter for an earlier edition.

## DISCLAIMER

The findings and conclusions in this chapter are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

## REFERENCES

- Adams, M.A. (1995). Mechanical testing of the spine: An appraisal of methodology, results, and conclusions. *Spine*, 20, 2151–2156.
- Adams, M.A. and Hutton, W.C. (1981). The effect of posture on the strength of the lumbar spine. *Engineering in Medicine*, 10, 199–202.
- Aspden, R.M. (1987). Intra-abdominal pressure and its role in spinal mechanics. *Clinical Biomechanics*, 2, 168–174.
- Ayoub, M.M. and El Bassoussi, M.M. (1978). Dynamic biomechanical model for sagittal plane lifting activities. In *Safety in Manual Materials Handling*, ed. C.G. Drury (Cincinnati, OH: U.S. Department of Health, Education and Welfare), pp. 88–95.
- Barford, N.C. (1985). *Experimental Measurements: Precision, Error and Truth*, 2nd edn. (New York: John Wiley).
- Bean, J.C., Chaffin, D.B. and Schultz, A.B. (1988). Biomechanical model calculation of muscle contraction forces: A double linear programming method. *Journal of Biomechanics*, 21, 59–66.
- Brinckmann, P. (1986). Injury of the annulus fibrosus and disc protrusions. An in vitro investigation on human lumbar discs. *Spine*, 11, 149–153.
- Brinckmann, P., Johannleueling, N., Hilweg, D. and Biggemann, M. (1987). Fatigue fracture of human lumbar vertebrae. *Clinical Biomechanics*, 2, 94–96.
- Chaffin, D.B. (1988). A biomechanical strength model for use in industry. *Applied Industrial Hygiene*, 3, 79–86.
- Chaffin, D.B. (1999). Static biomechanical modelling in manual lifting. In *The Occupational Ergonomics Handbook*, eds W. Karwowski and W.S. Marras (Boca Raton, FL: CRC Press).
- Chaffin, D.B. and Andersson, G.J. (1984). *Occupational Biomechanics* (New York: Wiley-Interscience).
- Chaffin, D.B., Andersson, G.J. and Bernard, B.J. (1999). *Occupational Biomechanics*, 3rd edn. (New York: Wiley-Interscience).
- Chaffin, D.B., Redfern, M.S., Erig, M. and Goldstein, S.A. (1990). Lumbar muscle size and locations from CT scans of 96 women of age 40 to 63 years. *Clinical Biomechanics*, 5, 9–16.
- Chang, C.C., Hsiang, S., Dempsey, P.G. and McGorry, R.W. (2003). A computerised video-based biomechanical analysis tool for lifting tasks: Model development and software design. *International Journal of Industrial Ergonomics*, 32(4), 239–250.
- Cholewicki, J., McGill, S.M. and Norman, R.W. (1995). Comparison of muscle forces and joint load from an optimisation and EMG assisted lumbar spine model: Towards development of a hybrid approach. *Journal of Biomechanics*, 28, 321–331.

- Daggfeldt, K. and Thorstensson, A. (2003). The mechanics of torque production about the lumbar spine. *Journal of Biomechanics*, 36, 815–825.
- Danz, M.E. and Ayoub, M.M. (1992). The effects of speed, frequency, and load on measured hand forces for a floor to knuckle lifting task. *Ergonomics*, 35, 833–843.
- Davis, P.R. and Stubbs, D.A. (1977). Safe levels of manual forces for young males. *Applied Ergonomics*, 8, 141–150.
- Dempsey, P.G. (1998). A critical review of biomechanical, epidemiological, physiological and psychophysical criteria for designing manual materials handling tasks. *Ergonomics*, 41, 73–88.
- Drury, C.G., Roberts, D.P., Hansgen, R. and Bayman, J.P. (1983). Evaluation of a palletising aid. *Applied Ergonomics*, 14, 242–246.
- Dutta, T. (2012). Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied Ergonomics*, 43(4), 645–649.
- Ferguson, S.A., Marras, W.S. and Burr, D. (2005). Workplace design guidelines for asymptomatic vs. low-back-injured workers. *Applied Ergonomics*, 36, 85–95.
- Fethke, N.B., Gant, L.C. and Gerr, F. (2011). Comparison of biomechanical loading during use of conventional stud welding equipment and an alternate system. *Applied Ergonomics*, 42(5), 725–734.
- Fong, D.T.-P. and Chan, Y.-Y. (2010). The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review. *Sensors*, 10, 11556–11565.
- Gagnon, M., Sicard, C. and Sirois, J.P. (1986). Evaluation of forces on the lumbo-sacral joint and assessment of work and energy transfers in nursing aides lifting patients. *Ergonomics*, 29, 407–421.
- Garg, A., Chaffin, D.B. and Freivalds, A. (1982). Biomechanical stresses from manual load lifting: A static vs dynamic evaluation. *HE Transactions*, 14, 272–281.
- Grieve, D.W. and Pheasant, S.T. (1982). Biomechanics. In *The Body at Work: Biological Ergonomics*, ed. W.T. Singleton (Cambridge: Cambridge University Press), pp. 71–161.
- Hsiang, S.M., Brogmus, G.E., Martin, S.E. and Bezverkhny, I.B. (1998). Video based lifting technique coding system. *Ergonomics*, 41, 239–256.
- Hsiang, S.M., Chang, C.C. and McGorry, R.W. (1999). Development of a set of equations describing joint trajectories during para-sagittal lifting. *Journal of Biomechanics*, 32, 871–876.
- Hughes, R.E. and Chaffin, D.B. (1995). The effect of strict muscle stress limits on abdominal muscle force predictions for combined torsion and extension loadings. *Journal of Biomechanics*, 28, 527–533.
- Hutton, W.C. and Adams, M.A. (1982). Can the lumbar spine be crushed in heavy lifting? *Spine*, 7, 586–590.
- Jäger, M., Jordan, C., Luttmann, A. and Laurig, W. (2000). Evaluation and assessment of lumbar load during total shifts for occupational materials handling jobs within the Dortmund Lumbar Load Study – DOLLY. *International Journal of Industrial Ergonomics*, 25, 553–571.
- Jäger, M. and Luttmann, A. (1986). Biomechanical model calculations of spinal stress for different working postures in various workload situations. In *The Ergonomics of Working Postures*, eds. E.N. Corlett, J.R. Wilson and I. Manenica (London, U.K.: Taylor & Francis Group), pp. 144–154.
- Jäger, M. and Luttmann, A. (1997). Assessment of low-back load during manual materials handling. In *Proceedings of the 13th Triennial Congress of the International Ergonomics Association*, Vol. 4, eds. P. Seppälä, T. Luopajarvi, C.H. Nygård and M. Mattila (Helsinki, Finland: Finnish Institute of Occupational Health), pp. 171–173.
- Jäger, M. and Luttmann, A. (1999). Critical survey on the biomechanical criterion in the NIOSH method for the design and evaluation of manual lifting tasks. *International Journal of Industrial Ergonomics*, 23, 331–337.
- Joint Motion: Method of Measuring and Recording. Published by the American Academy of Orthopaedic Surgeons, 1965. Reprinted by the British Orthopaedic Association, 1966. Edinburgh and London: E. & S. Livingstone Ltd.
- Jones, D.F. (1983). Back injury research: Have we overlooked something? *Journal of Safety Research*, 14, 53–64.
- Krag, M.H., Gilbertson, L. and Pope, M.H. (1985). Intra-abdominal and intra-thoracic pressure effects upon load bearing of the spine. In *31st Annual Meeting Orthopedic Research Society* (Las Vegas, NV).
- Kumar, S. (1988). Moment arms of spinal musculature determined from CT scans. *Clinical Biomechanics*, 3, 137–144.
- Kumar, S. (1990). Cumulative load as a risk factor for back pain. *Spine*, 15, 1311–1316.
- Leskinen, T.P.J. and Troup, J.D.G. (1984). The effect of intra-abdominal pressure on lumbosacral compression when lifting. In *Computer-Aided Biomedical Imaging and Graphics Physiological Measurement and Control: Proceedings* (Aberdeen, U.K.: PMCS), p. 4.

- Marras, W.S. and Granata, K.P. (1997). The development of an EMG-assisted model to assess spine loading during whole-body free-dynamic lifting. *Journal of Electromyographic Kinesiology*, 7, 259–268.
- McGill, S.M. (1999). Dynamic low back models: Theory and relevance in assisting the ergonomist to reduce the risk of low back injury. In *The Occupational Ergonomics Handbook*, eds. W. Karwowski and W.S. Marras (Boca Raton, FL: CRC Press).
- McGill, S.M. and Norman, R.W. (1985). Dynamically and statically determined low-back moments during lifting. *Journal of Biomechanics*, 18, 877–885.
- McGorry, R.W., Chang, C.C., Teare, P.R. and Dempsey, P.G. (2002). The flexible handheld ergonomics evaluation tool. *Ergonomics in Design*, 10(4), 5–11.
- Nachemson, A.L., Andersson, G.B.J. and Schultz, A.B. (1986). Valsalva maneuver biomechanics: Effects on lumbar trunk loads of elevated intra-abdominal pressures. *Spine*, 11, 476–479.
- Nachemson, A.L. and Elfström, G. (1970). Intravital dynamic pressure measurements in lumbar discs. *Scandinavian Journal of Rehabilitation Medicine. Supplement*, 1, 1–40.
- Nachemson, A.L. and Morris, J. (1964). In vivo measurements of intradiscal pressure. *Journal of Bone and Joint Surgery*, 46A, 1077–1092.
- Nemeth, G. and Ohlsen, H. (1986). Moment arm lengths of trunk muscles to the lumbosacral joint obtained in vivo with computed tomography. *Spine*, 11, 158–160.
- NIOSH (National Institute for Occupational Safety and Health). (1981). *A Work Practices Guide for Manual Lifting* (Cincinnati, OH: DHHS (NIOSH)), Publication No. 81-122.
- Norman, R.W. and McGill, S.M. (1999). Selection of 2-D and 3-D biomechanical spine models: Issues for consideration by the ergonomist. In *The Occupational Ergonomics Handbook*, eds. W. Karwowski and W.S. Marras (Boca Raton, FL: CRC Press).
- Norman, R.W., Wells, R., Neumann, P., Frank, J., Shannon, H. and Kerr, M., OUBPS Group. (1998). A comparison of peak vs cumulative physical work exposure risk factors for the reporting of low back pain in the automotive industry. *Clinical Biomechanics*, 13, 561–573.
- Nussbaum, M.A., Chaffin, D.B. and Martin, B.J. (1995). A back-propagation neural network model of lumbar muscle recruitment during moderate static exertions. *Journal of Biomechanics*, 28(9), 1015–1024.
- Perez, M.A. and Nussbaum, M.A. (2003). Principle components analysis as an evaluation and classification tool for lower torso sEMG data. *Journal of Biomechanics*, 36, 1225–1229.
- Pheasant, S. (1986). *Bodyspace: Anthropometry, Ergonomics and Design* (London, U.K.: Taylor & Francis Group).
- Schultz, A.B. and Andersson, G.B.J. (1981). Analysis of loads on the lumbar spine. *Spine*, 6, 76–82.
- Schultz, A.B., Andersson, G.B.J., Örtengren, R., Haderspeck, K. and Nachemson, A. (1982). Loads on the lumbar spine. *Journal of Bone and Joint Surgery*, 64A, 713–720.
- Schultz, A., Haderspeck, K., Warwick, D. and Portillo, D. (1983). The use of lumbar trunk muscles in isometric performance of mechanically complex standing tasks. *Journal of Orthopaedic Research*, 1, 77–91.
- Stokes, I.A.F. and Moreland, M.S. (1987). Measurement of the shape of the surface of the back in patients with scoliosis. *Journal of Bone and Joint Surgery*, 69A, 203–211.
- Svensson, O.K. (1987). On quantification of muscular load during standing work: A biomechanical study. Dissertation from the Kinesiology Research Group, Department of Anatomy, Karolinska Institute, Stockholm, Sweden.
- Thewlis, D., Bishop, C., Daniell, N. and Paul, G. (2013). Next-generation low-cost motion capture systems can provide comparable spatial accuracy to high-end systems. *Journal of Applied Biomechanics*, 29, 112–117.
- Tracy, M.F. (1988). Strength and posture guidelines: A biomechanical approach. PhD thesis, University of Nottingham, Nottingham, U.K.
- Tracy, M.F., Gibson, M.J., Szypryt, E.P., Rutherford, A. and Corlett, E.N. (1989). The geometry of the lumbar spine determined by magnetic resonance imaging. *Spine*, 14, 186–193.
- Troup, J.D.G., Foreman, T.K., Baxter, C.E. and Brown, D. (1987). The perception of back pain and the role of psychophysical tests of lifting capacity. *Spine*, 12, 645–657.
- Tsuang, Y.H., Schipplein, O.D., Trafimow, J.H. and Andersson, G.B.J. (1992). Influence of body segment dynamics on loads at the lumbar spine during lifting. *Ergonomics*, 35, 437–444.
- Vignais, N., Miezel, M., Bleser, G., Mura, K. and Gorecky, D. (2013). Innovative system for real-time ergonomic feedback in industrial manufacturing. *Applied Ergonomics*, 44, 566–574.
- Waters, T.R., Putz-Anderson, V., Garg, A. and Fine, L.J. (1993). Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36, 749–776.
- Xu, X., Chang, C.C., Faber, G.S., Kingma, I. and Dennerlein, J.T. (2012). Estimation of 3-D peak L5/S1 joint moment during asymmetric lifting tasks with cubic spline interpolation of segment Euler angles. *Applied Ergonomics*, 43, 115–120.



---

# 18 Definition and Measurement of Human Workload

*Sarah Sharples and Ted Megaw\**

## CONTENTS

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| Introduction.....                                                   | 516 |
| Underlying Psychological Processes .....                            | 519 |
| Classification of Mental Workload Techniques and Measures .....     | 524 |
| Criteria for MWL Techniques and Measures.....                       | 524 |
| Review of Some of the More Popular MWL Measurement Techniques ..... | 527 |
| Analytic Techniques.....                                            | 527 |
| Comparability Analysis .....                                        | 527 |
| Mathematical Models.....                                            | 528 |
| Expert Opinion .....                                                | 528 |
| Task Analytic Methods.....                                          | 528 |
| Simulation Models .....                                             | 529 |
| Empirical Techniques.....                                           | 530 |
| Primary Task Performance .....                                      | 530 |
| Secondary Task Techniques.....                                      | 530 |
| Psychophysiological Techniques.....                                 | 533 |
| Cardiac Activity.....                                               | 534 |
| Brain Activity .....                                                | 535 |
| Electrodermal Activity .....                                        | 536 |
| Eye Function .....                                                  | 536 |
| Body Fluid Analysis .....                                           | 536 |
| Muscle and Movement Analysis .....                                  | 537 |
| Subjective/Operator Opinion Techniques .....                        | 537 |
| Single-Dimensional Scales.....                                      | 537 |
| Multidimensional Scales .....                                       | 538 |
| Relative Judgements.....                                            | 540 |
| Instantaneous Judgements.....                                       | 540 |
| Comparison between Rating Scales .....                              | 541 |
| Interviews .....                                                    | 542 |
| Observations .....                                                  | 542 |
| Validity of Subjective Measures .....                               | 542 |
| Conclusions.....                                                    | 543 |
| References.....                                                     | 544 |

---

\* Note, this chapter has been written by Sharples, based on the previous version by Megaw, therefore throughout the chapter 'I' refers to Sharples and 'we' refers to the combined views of both authors.

## INTRODUCTION

As jobs in the twenty-first century continue to involve the use of an increasing range of technologies and systems, tasks are increasingly dominated by mental rather than physical task components. Modern-day cars now routinely include features such as cruise control and parking sensors that *reduce the load* on the driver, as well as introducing complex communication and entertainment technologies to improve the passenger's comfort and experience. As I work on this chapter, there are seven separate electronic devices on my desk that are supporting my task, ranging from a tablet computer to a desktop clock. Similarly, many high-risk or safety-critical jobs, such as pilots, train drivers, medical personnel and process control operators, are characterised by their mental rather than physical demands.

Intuitively, it is reasonable to assume that it is essential that the work demands are within the capabilities of those performing jobs to ensure acceptable system performance. Critically, it is also important that the spare capacity of an operator is sufficient for them to be able to respond in the case of an unexpected event or emergency situation. The need to be able to understand current workloads of operators and predict their future capacity to cope with additional work has led over the last 50 years to the increasing effort to develop reliable, valid and practical measures of what has traditionally been termed *mental workload* (MWL). Within this chapter the terms MWL and workload are both used; as discussed later, the definition and scope of the term MWL is interpreted differently by different authors and in different tools. We present the view that the concept of workload is one that is predominately concerned with cognitive (i.e. mental activities) but which takes place within a context that is social and interacts with and should not ignore the physical elements of tasks and work.

There has been much discussion surrounding the question of providing an acceptable definition of MWL in order to structure the measurement process. As with other areas of Ergonomics/Human Factors (E/HF), for example visual fatigue and situation awareness (see Chapter 19), this has proven difficult. This can be appreciated if one considers the wide array of definitions of MWL offered by the various contributors to the first significant collection of papers devoted to the subject (Moray, 1979). And things have not got that much better! In 1989, Linton et al. stated, 'The simple fact of the matter is that nobody seems to know what workload is. Numerous definitions have been proposed, and many of them seem complete and intuitively *right*. Nevertheless, current definitions of workload all fail to stand the test of widespread acceptance or quantitative validation' [p. 22]. For example, Young and Stanton (2002b) propose a definition of MWL as representing 'the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support and past experience' but themselves identify limitations of this definition including its focus on attention as the underlying cognitive mechanism of interest, and the explicit role of subjectivity and internal goals. Wickens (2008) returns to Moray's (1979) work and notes the key characteristic of workload as 'the demand imposed by tasks on the human's limited resources, whether considered single or multiple'. The culmination of this uncertainty is reflected in the contents of the international standard (ISO 10075, 2000) entitled *Ergonomic Principles Related to Mental Work-Load*, in particular, Part 1: General Terms and Definitions. This standards document offers no definition of MWL, but rather (and unsatisfactorily) defines terms such as mental stress and mental strain in the context of the simple engineering or stimulus-based approach to stress. Within this approach, stress(ors) is defined in terms of the influences impinging upon a person, and strain as the effects of those stressors on the individual (Cox, 1978). Interestingly, this definition was originally published in the 1991 version of the standard and was not updated in the 2000 revision.

As noted earlier, even the phrase *mental workload* can be contentious or interpreted differently by different researchers. In much of the first author's (Sharples) work, we have tended to use the term *workload* rather than MWL to denote the interacting physical and cognitive elements of

workload that can not only be seen in more *manual* jobs, such as a supermarket checkout operator for example, as well as the ability to observe physical interactions and use these observations to support inferences about cognitive effort. In the field of human-computer interaction, many authors who are now interested in understanding the demands placed upon users of interactive systems have (perhaps inappropriately) adopted the term *cognitive load*. While this term is one that intuitively suggests a holistic consideration of the cognitive aspects of a work task, much of the original theory on which it is based (e.g. see Chandler and Sweller, 1991; Paas et al., 2003) is focused on laboratory-based problem-solving tasks and instructional methods and very much takes a working memory-driven cognitive psychology approach to the issue of load. Human factors researchers are inevitably working with real-world tasks or jobs where expertise, memory, attention, situation awareness and social and organisational factors all combine to contribute to the individual's experience of workload and thus the concept of MWL needs to reflect this real-world complexity.

Rather than pursue a succinct definition of MWL, it is more profitable to provide a simple framework to help understand the implications for the actual measurement process of MWL. The most important feature of the framework is its dynamic nature. This framework is based upon the version published in the 2005 edition of this book, but also informed by the work of Pickup et al. (2005b). As shown in Figure 18.1, there are three main components to the framework: physical and cognitive task demands; operator workload/effort; and performance. As will be described in the main part of this chapter, measurements of these components, as well as the external and internal influences on workload, can be obtained to a greater or lesser effect. However, the essence of MWL and its measurement is reflected in the relationships between the three components.

The *physical and cognitive task demands* reflect the characteristics of the task or tasks undertaken by a person and thus imposed upon a person. If we are to effectively monitor MWL over time or predict operator workload in different work circumstances, it is reasonable to assume that an objective and quantifiable measure of work demand is desirable. It is important that both physical and cognitive elements of demand are captured as these may both influence an individual's reports of their perceptions of the workload they are experiencing. Attempting to quantify physical and cognitive task demands, however, is extremely challenging except in the case of very simple tasks. Similarly it is important to acknowledge that the externally measurable demand on the individual may be different from the demand perceived by the individual, thus emphasising the importance of not only capturing the externally imposed demands but also how workload is perceived. As will be seen later in this text, different methods take different account of the physical element of work when measuring MWL. The inclusion of the physical element within this framework recognises that the interaction between physical and cognitive exists, and that if we are attempting to capture cognitive elements of workload, it is likely that some element of physical work will also be present and may affect the experience of the individual and the data captured in the measurement of workload.

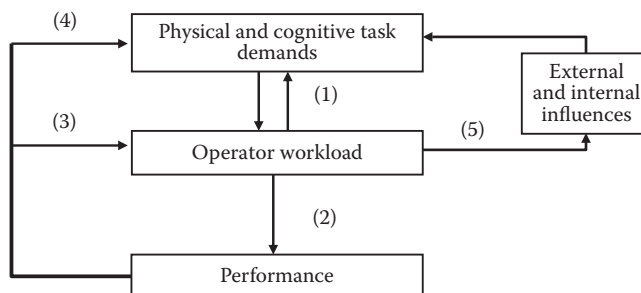


FIGURE 18.1 A framework for MWL definition and evaluation.

*Operator workload* is conceived in terms of the operator performing the task and, if we consider the engineering approach to stress research, is equivalent to measures of operator strain or effort. A majority of MWL measurement is concerned with measures of this kind and focuses on capturing the operator's experience of the work via subjective report but can also include inferences of effort from behaviour indices and interaction logs.

*Performance* is self-explanatory, and most frequently is described in terms of speed and errors (and, of course, therefore is subject to many challenges in data interpretation resulting from the classic time/error trade-off). However, the complexity of the measurement problem is only appreciated when one realises that there are few, if any, linear or even monotonic relationships between measures of these three MWL components of demand, workload and performance. That is to say, contrary to what might be expected, as the task demands increase there is not necessarily an increase in operator workload or decrease in task performance. An indication of why such simple relationships do not exist emerges when one considers the various relationships as numbered in Figure 18.1:

1. Operator workload or effort is not simply a function of task demands but is influenced by how the task is perceived by the operator, sometimes referred to as cognitive appraisal. Pickup et al. (2005b) state that effort can be seen as a consequence of demand created by loading factors, represented in Figure 18.1 by a combination of the physical and cognitive task demands and external and internal influences. Workload is influenced by intrinsic factors (noted as *external and internal influences*) such as levels of operator skill, amount of practice and training and attitudes towards or enthusiasm for the task (i.e. we may work harder and perceive the task as less demanding if it is one which we enjoy or which we can see will reward us when complete). The intrinsic factors will also determine the strategies that the operator adopts, so as well as the workload being perceived as being lower, the actual workload is effectively reduced if efficient strategies are adopted. Conversely, low arousal levels and fatigue can cause the adoption of less effective strategies and lead to lower actual and perceived workload. These factors indicate that measures of operator workload will not necessarily reflect objective measures of task demands but can be a confounded mix of perceptions, attitudes, strategies and actions. For example, just because a learner driver records a high level of workload does not imply that the driving task is imposing too many demands and should, therefore, be modified, but rather that there is need for further practice and training.
2. Although it is often assumed that there is a close relationship between operator workload and task performance, with high operator workload being associated with relatively poorer performance, dissociations between workload and performance frequently occur. Even when a sensitive performance measure (such as a detailed task interaction logging tool) is available, it is likely to be unable to detect the phenomenon where an individual is working harder (i.e. incurring additional cognitive effort) to maintain a consistent level of performance.
3. Operators monitor their own performance, both unconsciously and explicitly, as well as being provided with numerous sources of feedback. This may change the way they perceive the task, alter their performance strategies and have motivational consequences, and hence modify their workload. This is analogous to the cognitive appraisal feedback loop found in the transactional model of stress (Cox and Mackay, 1981).
4. Performance outcomes can modify the tasks themselves so that task demands are altered. For example, in the case of a process control operator diagnosing a complex fault, it is possible that by following some inappropriate diagnostic strategy (that may have occurred due to high demands on the operator), the subsequent task demands are increased. This could lead to an exponential effect and has been observed in the context of air traffic control (ATC) (see Sharples et al., 2012) where multiple factors combine to lead to a rapid, non-graceful, decline in performance (which we term the *precipice of performance*).

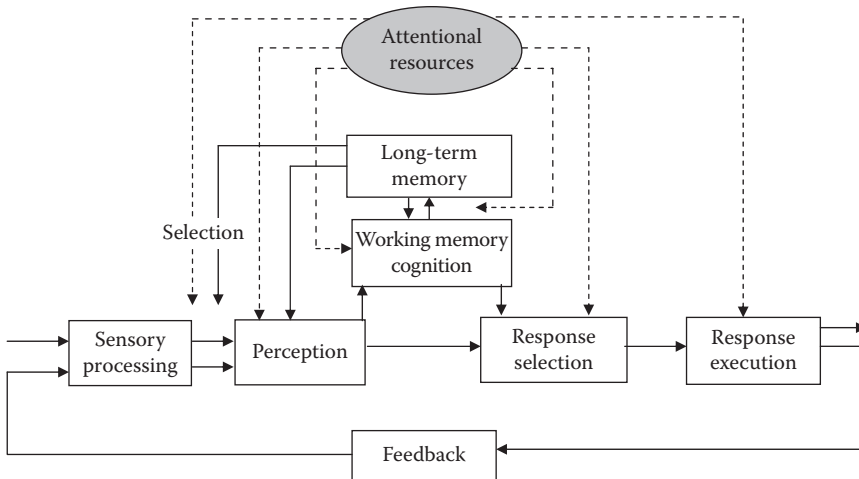
5. Task and jobs happen in a work context. External factors, such as job type, organisational and safety culture and social support within the work environment, will influence operator behaviour and experiences or perceptions of workload. For example, an operator working in a supportive team environment with a strong safety culture will be more confident to ask for help when experiencing high workload and may request support from a supervisor more quickly, thus maintaining their work demands at a manageable level and avoiding the descent over the *precipice of performance*. In addition to these external factors, internal factors, such as operator skill and motivation (as noted in point 1), will influence both absolute perception of workload and selection of behavioural strategies that will have an indirect effect on experienced workload via adjusted task demands.

These last three interactions also highlight the fact that there are likely to be temporal fluctuations in MWL levels while performing essentially the same job. However, these fluctuations are not easy to assess and demand longitudinal measures if they are to be captured at all.

The external influences noted in Figure 18.1 are often discussed with reference to an inverted-U relationship relating task performance to levels of physiological arousal, whereby performance is relatively poor under conditions of low (under arousal) or high (over arousal) levels. Physiological arousal levels are a function of both task demands and the impinging external stressors. By applying the contentious Yerkes–Dodson law (Yerkes and Dodson, 1908) to the inverted-U relationship, a number of predictions can be made about the influences of external stressors on primary task performance. Vigilance-related tasks, characteristic of many monitoring or signal detection tasks, might be enhanced by the addition of external stimuli such as music-while-you-work or white noise (Smith, 1989). Such tasks pose low workload demands and hence yield low levels of intrinsic arousal. The role of the external stimuli would, therefore, be to increase arousal levels and thus improve performance. It is important to take into account individual characteristics and the nature of such external influences however. If, for example, we consider someone completing a CCTV monitoring task in a context where there is a low frequency of incidents, we may classify this task as having a low demand. The background noise of their environment may have the same, beneficial effect as *white noise*, but, if something meaningful to the operator is heard, such as the operator's own name (as in the classic *Cocktail Party Effect*; Cherry, 1953) this will distract the operator and potentially have a negative effect on task performance. Taking an example at the other extreme, although performance in respect of a fairly high demanding task might be satisfactory under certain training conditions, one might find that such performance levels do not transfer to real operational conditions. This is because the increased arousal resulting from stressors such as personal fear and a range of environmental factors can lead to conditions of over arousal. Conversely, performance in a simulated context may be of lower quality compared with the real world due to motivational factors. Thus, external stimuli may cause both performance enhancement and degradation, depending on the task demands.

## UNDERLYING PSYCHOLOGICAL PROCESSES

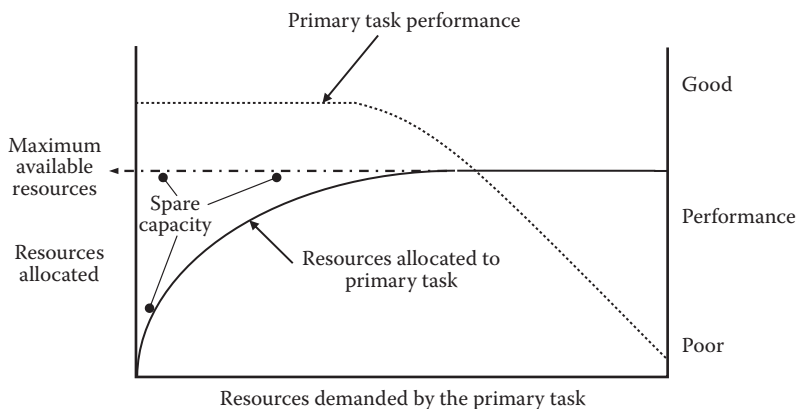
Much of our current understanding of MWL has been influenced by what were originally described as models of information processing. According to these early models (Welford, 1968), humans are characterised by possessing a limited amount of information processing capacity such that performance deteriorates if task demands exceed this capacity. In the early models, this concept of limited capacity remained fairly ill-defined, though attempts to measure the capacity through the application of information theory were prevalent (e.g. as reflected in the work of Hick [1952] and Fitts [1954]). A number of key refinements were then made by authors such as Kahneman (1973) and Wickens (Wickens 2008; Wickens and Hollands, 1999) which have important implications for the definition and measurement of MWL. Most significantly, the term *limited processing capacity* was replaced by the term *attentional resources*. These resources have to be shared between a number



**FIGURE 18.2** A general model of human information processing. (Adapted from Wickens, C.D. and Hollands, J.G., *Engineering Psychology and Human Performance*, 3rd edn., Prentice Hall, Upper Saddle River, NJ, 1999.)

of basic psychological processes such as perception, working memory and response execution, depending on the task demands. This role of attentional resources in human information processing is illustrated in the general model proposed by Wickens, which is shown in Figure 18.2.

MWL can then be conceived in terms of the relationship between the resources supplied to these various processes and the task demands in the manner depicted in Figure 18.3. The vertical axis on the left indicates the resources being used by the task and denotes the maximum level of resources available; the vertical axis on the right indicates the performance on the primary task (denoted by the dotted line on the graph). On the left of the figure is a region (described by Wickens et al. 2013 as the *reserve capacity region*) where task performance is satisfactory because the resources available are in excess of the task demands, and there is spare capacity or there are spare attentional resources. This region reflects conditions of relatively low-task demands where workload is inversely related to the amount of spare resource capacity. This therefore predicts that measures of



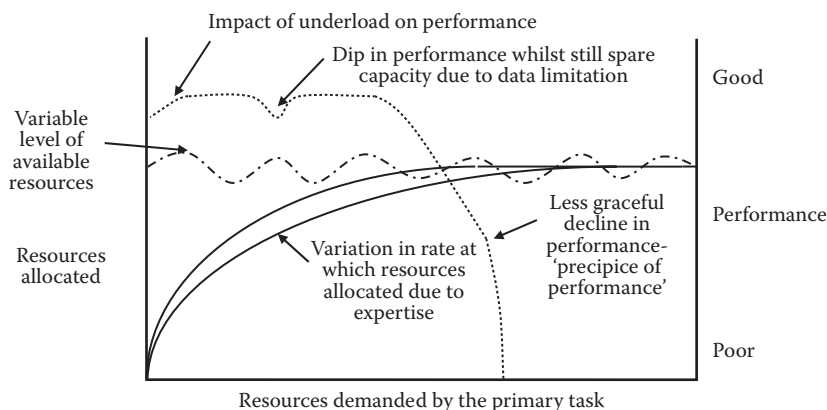
**FIGURE 18.3** The relationship between the resources allocated to the primary task and the resources demanded by the primary task (—), and the relationship between primary task performance and the resources demanded by the primary task (····). (Adapted from Wickens, C.D., Hollands, J.G., Banbury, S. and Parasuraman, R. (2013). *Engineering Psychology and Human Performance*, 4th edn., Boston, MA: Pearson.)



performance on the main task being undertaken – the primary task (particularly if they are summative measures of task completion, as opposed to descriptive or analytical measures of performance strategies) will not be sensitive to workload changes in this region. However, measures on a secondary task (a supplementary task introduced in addition to the primary, or main, task) should reflect the amount of spare capacity available. Moving to the right of the figure, one enters a region where insufficient resources are available to meet the task demands because the limit in available resources has been reached. This is a region where task demands are relatively high and where workload is represented as being inversely related to primary task performance. Thus in this region, a measure of primary task performance should indeed allow inference of workload.

While the relationship between allocated resources, demanded resources and primary task performance as presented in Figure 18.3 is a very useful hypothetical model of the limited resources approach, it raises a number of issues. These are noted later, and Figure 18.4 illustrates the potential impact of some of these issues on the graphical representation of the model:

1. *External and internal influences:* The paradigm adopted in the model, if taken in conjunction with the adoption of attentional resources as the primary type of resources being considered within MWL, implies that the amount of maximum available resources is a constant. This may or may not be the case, and to a certain extent is dependent on the extent to which the model is taken as a literal representation of what is going on within the individual's cognitive resources as opposed to a useful tool to explain the relationship between the different factors of resources, performance and work demands (I favour the latter approach!). If we consider the situation where a person is under a stressful situation (e.g. a junior doctor encounters a tricky diagnostic case while on-call in a hospital in which she has only been working for a few weeks), then, even if the external demands placed upon an individual are at a manageable level, the stress associated with the situation may have a negative influence on perceived workload and performance (time taken to correctly diagnose may be longer, or an incorrect diagnosis may be made). The model in Figure 18.3 does not capture this scenario well. We could consider the intrinsic element of stress to be an additional factor that is *taking up* some of the spare resources or we could (perhaps more usefully) consider the maximum level of resources to have some degree of variability (and in the example here, be reduced). Similarly, if we approach a task with a clear mind and high level of motivation, we could consider our maximum level of resources to be increased. This has some analogies with the Malleable Attentional



**FIGURE 18.4** Variations on the relationship between the resources allocated to the primary task and the resources demanded by the primary task (—), and the relationship between primary task performance and the resources demanded by the primary task (···).

Resources Theory (Young and Stanton, 2002a), although this theory focuses on the role of task demands on attentional resources, rather than external or internal factors, as in the example given here.

2. *Impact of expertise and automaticity:* It has been identified (Schneider and Shiffrin, 1977; Treisman and Gellade, 1980) in visual search tasks that where elements were consistently mapped together in relation to stimulus characteristics and required responses, performance was not degraded by increasing task difficulty. The consistent mapping is usually achieved after an individual has received plenty of experience of the task characteristics. However, with inconsistent mapping, even after practice, performance decreases with increased task difficulty. In the former case, it can be concluded that the task does not therefore require the investment of attentional resources, while in the latter it does. Another example of automatic processes is the execution of skill-based motor performance, such as walking. After an extensive learning phase, walking is controlled in an open-loop rather than a closed-loop manner whereby a command is issued to the brain instructing a walking motor programme to be initiated. Of course, there are some peripheral closed-loop activities to overcome, for example unevenness in the walking surface, but these are carried out unconsciously. However, if the unevenness becomes too great, for example when walking across a rocky surface, then conscious closed-loop control, requiring attentional resources, occurs. The role of expertise could be represented in the model in Figure 18.3 by showing solid lines with different ascending gradients (see Figure 18.4), depending on the level of expertise of the individual (where an expert performer would use up resources more slowly, represented by a more shallow gradient), but it would need to be remembered that the relationship between mappings of stimulus and response will also have an impact on the extent to which expertise can reduce the resources demanded by the task.
3. *Underload:* As Young and Stanton (2002a) highlight, performance may degrade in circumstances of low-task resource demand as well as high demand. Whether this performance degradation is due to lowered arousal or reduced attentional resources, it has been demonstrated to occur and it is important to identify whether degraded performance is in fact due to maximum available resources having been exceeded rather than being due to underload. This can normally be achieved by measurement of task demand and subjective perception of workload as well as performance.
4. *Linearity of relationships and graceful performance degradation:* As mentioned earlier, the graphical representations in Figure 18.3 are hypothetical and it may well be the case that the increase in workload as task demands increase is not graceful. Our work in ATC (Sharples et al., 2012; Edwards et al., 2012) has confirmed the likelihood of the existence of a *precipice of performance* (as represented in Figure 18.4). In addition, the complexity of the transition from low to high workload represents complexity of the relationship between effort and demand (Pickup et al., 2005b). A small increase in both expended effort and imposed demand may represent the descent onto the precipice of performance (analogous to the last straw on the camel's back!) and may have a dramatic effect on the individual's primary task performance.
5. *Tasks, jobs and work context:* The role of external factors such as organisational culture and teamwork has already been highlighted, but it is important to clarify that many *jobs* consist of a, sometimes quite large, collection of *tasks* (Nichols et al., 2001). Traditional instruments for MWL measurement have focused on task completion, and the interacting nature of different tasks with different temporal patterns, using different sensory modalities and imposing different levels of difficulty are not represented within the model in Figure 18.3 (see later in this chapter for further discussion of handling multiple sensory modalities in workload).
6. *Task switching:* Following on from this, it is apparent that under several circumstances, depending on the specific underlying workload demands, certain tasks or subtasks cannot

be efficiently carried out in a parallel manner. In such cases, it is necessary, therefore, for operators to switch between tasks or subtasks. This process can be thought of as task management or task switching (see Monsell, 2003 for a useful description of cognitive elements of task switching). A critical aspect of the introduction of automation is to ensure operators adopt optimum task management strategies. A typical task where switching is important is the monitoring of multiple displays in police and process control rooms, hospital operating theatres and aircraft and other vehicle workplaces. How operators should allocate the time they spend observing each display was first discussed by Senders (1964). While task switching is an essential component of efficient performance, it should be noted that switching does impose its own demands and these demands or *switch costs* (Rubenstein et al., 2001; Wylie and Allport, 2000) will vary depending on specific characteristics of the different tasks. Further problems may arise from the extra perceived and actual demands imposed by task switching, resulting in operators being reluctant to switch and spend too much time on low-priority tasks at the expense of higher-priority tasks, a process referred to as *cognitive lock-up*. The 3D model of MWL developed by the TNO Human Factors Group (Neerinx, 2003) specifically includes task switching as one of the MWL dimensions, the other two being the percentage of available time needed to carry out the task and the level of processing demands.

7. *Data-limited versus resource-limited performance*: Norman and Bobrow (1975), in a somewhat hypothetical paper, discuss the relationships that may exist between attentional resources invested and performance. They postulated that some aspects of performance are not determined by the amount of resources invested because the performance is limited by the quality of the data (data-limited performance), while other aspects of performance are improved by increasing the amount of resources invested (resource-limited performance). Examples of data-limited aspects of task performance are visual detection and memory retrieval. In the former case, performance is limited by the discriminability or conspicuity of the target and, in the latter case, by the quality of the memory representation. This impact of data-limited performance is represented in Figure 18.4 by the inclusion of a *dip* in performance while still within the left-hand section of the graph, where spare capacity remains available. Naturally, many tasks will include both data- and resource-limited components – with the increase in use of large data sets for monitoring, planning and control tasks (reflected in the *big data* challenge; Mayer-Schonberger and Cukier, 2013), it is important that methods to measure MWL are able to capture moments where the representation or exclusion of data has a negative impact on performance, and that the methods are able to distinguish this type of lowered performance from that which results from task demands that exceed resources available.
8. *Single or multiple resources*: Finally, but perhaps most importantly of all, many tasks we undertake are made up of a number of subtasks, all of which may impose different demands on different sensory, perceptual and response modalities. This is best represented by Wickens' (2002) multiple resource theory. This theory emphasises the distinct resource sets associated with, in particular, the visual and auditory modalities, and demonstrates that resources are allocated according to modalities, stages of processing, codes of processing and response types (Wickens, 2008). Wickens himself, in his 2008 paper reflecting on the rationale for multiple resources theory, identifies some limitations in this approach (notably the omission of a haptic/tactile dimension and the need to understand what *drives* the allocation of resources). Wickens also notes that multiple resources theory is distinct from workload, mentioning that they are often confused, and that the contribution of multiple resources theory is not only in the consideration of demand (traditionally the focus of workload) but also in understanding the impact of resource overlap and allocation policy. According to the multiple resource model, the extent to which tasks or task elements can be executed in parallel will depend on the extent they compete for the same specific

attentional resources. Chapter 22, in discussion of neuroergonomics methods, highlights a number of studies that provide evidence in support of this multiple resources model. This is important when considering the measurement of the spare capacity, as noted in Figure 18.3, and is discussed further in the section on secondary task techniques of MWL assessment.

From the aforementioned discussion, it is not difficult to realise why it has been challenging to arrive at an acceptable definition of MWL. Are we trying to use our knowledge, although incomplete, of underlying models of psychological processes to derive definitions and measures of MWL, or are we using MWL measures to investigate models of psychological processes? Or perhaps we are using the meaningful and intuitive notion of *workload* to provide tools that are useful for directly informing workplace design (in which case, is there a danger of ignoring the underlying psychological processes and misusing or misinterpreting workload measurement tools?).

## CLASSIFICATION OF MENTAL WORKLOAD TECHNIQUES AND MEASURES

The classification shown in Table 18.1 is based on the framework provided by Hill et al. (1987). Although probably not exhaustive, the list does include what continue to be the most important measures of MWL. The main division of the techniques is into analytic and empirical measures, the latter requiring operators to perform the task(s) under investigation unlike the former where no such participation is required. The analytic techniques are frequently used to provide predictive estimates of MWL and are often incorporated into the overall system design process.

### CRITERIA FOR MWL TECHNIQUES AND MEASURES

A number of attempts have been made to establish the appropriate criteria for MWL measurement techniques. Those proposed by previous authors (e.g. O'Donnell and Eggemeier, 1986; Wickens and Hollands, 1999) are combined with those discussed in Chapter 1 to produce the following list:

- *Validity*: Broadly defined as the extent to which a method measures what it purports to measure, in the context of workload, face validity, construct validity and concurrent/convergent validity are particularly important. Face validity is, perhaps somewhat surprisingly, a challenge, as the obvious way to make, for example, a subjective measure of workload *look* like it is measuring workload is to include the word *workload* in a prompting item (a crude example would be *Please state whether your workload is high or low*). Yet, as discussed earlier, if the community of E/HF specialists are unable to come to a consensus as to the meaning of workload, it is unreasonable to expect consistent interpretation of the same word by subjective response respondents. Construct validity refers to whether a method measures all aspects associated with that construct and ensures it does not measure concepts irrelevant to that construct. The consideration of MWL as a unidimensional or multidimensional concept has been hinted at in our discussion of multiple resources theory, and the data captured by the selected measure(s) will of course influence the extent to which different elements of workload are measured. Finally, concurrent validity is a particularly useful concept for workload that encourages us to compare results from different measures – if two or more carefully applied measures suggest the same trend in workload we can have increased confidence in our findings; if they do not then we need to understand why, from either a theoretical or methodological perspective.
- *Reliability*: Reliability refers to consistency in outputs from methods over repeated administrations. For workload measures, this can be affected both by the quality/nature of the

**TABLE 18.1**  
**Classification of MWL Measurement Techniques**

| Method Type                                            |                                                                                                              | Example                                                                                                                                                                                     |
|--------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Analytic techniques</i>                             |                                                                                                              |                                                                                                                                                                                             |
| Comparability analysis                                 |                                                                                                              | Early Comparability Analysis Tool                                                                                                                                                           |
| Mathematical models                                    |                                                                                                              | Keystroke Level Model                                                                                                                                                                       |
| Expert opinion                                         |                                                                                                              | Pro-SWAT                                                                                                                                                                                    |
| Task analytic methods                                  |                                                                                                              | Cognitive Task Analysis                                                                                                                                                                     |
| Simulation models                                      |                                                                                                              | Micro Saint                                                                                                                                                                                 |
| <i>Empirical techniques</i>                            |                                                                                                              |                                                                                                                                                                                             |
| Primary task performance                               | Time related                                                                                                 | Time, error, strategy description, allocation of attention                                                                                                                                  |
|                                                        | Error related                                                                                                |                                                                                                                                                                                             |
|                                                        | Indices (combined measures)                                                                                  |                                                                                                                                                                                             |
|                                                        | Strategy related                                                                                             |                                                                                                                                                                                             |
| Secondary (concurrent or dual) task techniques         | Loading task                                                                                                 | Tracking task, mental arithmetic, conversation, handwriting task                                                                                                                            |
|                                                        | Subsidiary task                                                                                              |                                                                                                                                                                                             |
|                                                        | Adaptive task                                                                                                |                                                                                                                                                                                             |
|                                                        | Embedded task                                                                                                |                                                                                                                                                                                             |
| <i>Physiological or psychophysiological techniques</i> |                                                                                                              | Cardiorespiratory activity                                                                                                                                                                  |
|                                                        |                                                                                                              | Heart rate, heart rate variability, blood pressure, respiratory function                                                                                                                    |
|                                                        |                                                                                                              | Brain activity                                                                                                                                                                              |
|                                                        |                                                                                                              | MRI/fMRI, diffusion tensor imaging (DTI), fNIRS, transcranial Doppler (TCD), EEG, cortical evoked response/ERPs, transcranial magnetic stimulation (TMS), direct cortical stimulation (DCS) |
|                                                        |                                                                                                              | Electrodermal activity                                                                                                                                                                      |
|                                                        |                                                                                                              | GSR                                                                                                                                                                                         |
|                                                        |                                                                                                              | Eye function                                                                                                                                                                                |
|                                                        |                                                                                                              | Visual occlusion, eye movements, eye blinks, pupil response                                                                                                                                 |
|                                                        |                                                                                                              | Body fluid analysis                                                                                                                                                                         |
|                                                        |                                                                                                              | Hormonal analysis, blood glucose levels                                                                                                                                                     |
|                                                        |                                                                                                              | Muscle and movement analysis                                                                                                                                                                |
|                                                        |                                                                                                              | EMG, gait analysis                                                                                                                                                                          |
| <i>Operator opinion/subjective techniques</i>          | Rating scales: single dimensional or multi-dimensional, absolute or relative, retrospective or instantaneous | NASA-TLX, SWAT, Cooper–Harper, Air Force Flight Test Center (AFFTC)                                                                                                                         |
|                                                        | Interviews                                                                                                   | Behavioural marker elicitation, critical decision method                                                                                                                                    |
|                                                        | Observations                                                                                                 | Activity analysis/structured observation                                                                                                                                                    |

measure and the expertise of the administrator of the method or tool. For example, we demonstrated that a framework that we had established for observation of workload in rail signallers (Sharples et al., 2011) had inter-rater reliability of over 75% (Balfe, 2010).

- **Generalisability:** In general, as stated in Chapter 1, we should normally strive to develop and use methods that allow our results to transfer between contexts. In the case of workload measurement, it appears that in many cases it is appropriate or necessary to adopt domain-specific methods (see Integrated Workload Scale [IWS] [Pickup et al., 2005a], developed for rail, and Cooper–Harper Scale [Cooper and Harper, 1969], developed for aviation, described later in this chapter). An advantage of these domain-specific tools is that they allow domain-specific language to be used either in descriptors for experts or wording in subjective scale items, and there can also be some assumptions about the way in which certain words or phrases might be interpreted by groups of people who have had specific training.

- *Sensitivity*: Any measure needs to be able to detect changes in task demands, required effort or performance. In addition to the *traditional* considerations of sensitivity, such as required number of points on a rating scale or time periods, it is also important, particularly in the case of physiological measures, to avoid *oversensitivity* where a difference in a measure may perhaps even be statistically significant, but not meaningful in terms of inferring an indication of increased or decreased effort on the part of the operator.
- *Interference*: The measurement technique should not interfere with or disrupt primary task performance. This particularly applies when there are safety risks, for example, in the case of performing process control tasks or driving tasks under real operational settings, but is also important in any context to ensure realism and naturalness of operator performance. Increases in the sophistication of simulation technology have meant that we can be more confident now of interpreting data from simulated tasks; similarly, reductions in size of computer and sensory peripherals have made the introduction of direct physiological measures or subjective measures displayed on mobile or tablet computers into the workplace more feasible and less disruptive than in the past.
- *Diagnosticity*: A tool should allow the cause of variation in MWL to be identified. This can often be achieved by applying a combination of primary task performance, analytical and subjective measures. In addition, in the context of multiple resources theory, having a method that allows us to distinguish between demands on/effort incurred by different resources could lead to improved system design by ensuring that specific resources are not over-demanded. Diagnosticity also relates to need to be able to distinguish between the difference described earlier between data-limited and resource-limited performance.
- *Selectivity*: Related to construct validity, the technique should provide measures that allow variations in MWL to be distinguished from possible confounding factors including physical workload and emotional stress. It is important to note that this does not mean that these other types of factors should not be factors, in fact the contrary argument applies – it is important to capture and *distinguish between* these different influences and experiences to increase confidence in interpretation of findings as being related to workload, rather than something else.
- *Granularity/bandwidth*: The technique should provide measures that allow fairly rapid variations in MWL to be tracked over time, without risk of the measures saturating. Increased computing power means that very high sampling rates of performance or physiological measures are now feasible – the management and filtering of these large data sets needs to be very carefully considered.
- *Feasibility of use*: The extent to which a tool should be *simple* is a matter of much debate, but it is important to consider who is going to implement the tool and who is going to analyse the data. If expertise is required in either of these areas, this needs to be clearly indicated and, where possible, appropriate training or instructions provided.
- *Acceptability and ethics*: We need to be confident that participants will understand and follow the instructions given and cooperate with measurement techniques applied. In many safety-critical industries, employees are used to being closely monitored, and so in fact may be quite comfortable with having detailed measurement of their work performance. The availability of low-intrusion tools such as portable eye trackers and *spy glasses* (where a small camera is embedded into a pair of spectacles worn by an individual) presents ethical challenges as, while the individual who is making the recording will be fully aware of the presence of the recording tools, those around him/her may not. This introduces a debate that is beyond the scope of this chapter, but is nevertheless important to note.



- *Resources*: Finally, time and financial resources determine the practicality of selection of techniques. In the case of workload, the key determining factor is often availability of the workplace participants themselves; in many cases, such as from my own experience in the medical and rail domains, it is only meaningful to apply some measures of workload with expert participants while they are embedded within their real work domain.

A broad indication of the extent to which the various measurement techniques comply with the criteria such as those presented earlier is provided by O'Donnell and Eggemeier (1986). The following review of MWL techniques will consider these criteria. In addition to the methodological criteria, there are other aspects of workload that should be considered when selecting methods. The first is whether the method assumes a *multidimensional* or *unidimensional* theory of workload. For example, a method may focus on capturing the different elements that influence demand on an individual, but may not attempt to make any inference about the different types of cognitive effects that those elements might have (i.e. a unidimensional approach); alternatively, the method may particularly aim to capture different elements of workload, whether they be related to modalities of demand or types of demand (e.g. time demand vs. task complexity) (i.e. a multidimensional approach). Secondly, a method may strive to obtain an *absolute* measure of workload, where the measure, which is normally quantitative, aims to produce an indicator of workload that can be interpreted in isolation and has some meaning in the real world or to the individual (e.g. it allows an inference as to whether the individual is able to cope or not. This type of approach can be very difficult to validate and requires extensive benchmarks for interpretation (such as exist, e.g., in the case of physiological measures such as heart rate), but perhaps represents the *Holy Grail* of workload measurement, if we are able to use absolute measurement to infer where an individual's experience of workload sits with respect to graphical representations of resources, such as that shown in Figure 18.3. However, many measures are most usefully applied when they are considered in *relative* terms, enabling comparison of workload in two or more different conditions. The third type of consideration is whether a measure is *subjective* (i.e. its value depends on some interpretation or reporting on the part of the experimenter or participant) or *objective* (i.e. the measure is taken directly from the work situation or participant). This is a different consideration to the final distinction – that of whether the way in which the data from workload measures are reported – as either *qualitative* or *quantitative*.

## REVIEW OF SOME OF THE MORE POPULAR MWL MEASUREMENT TECHNIQUES

Clearly, it is not possible to cover all the techniques in this chapter. We therefore concentrate on the more popular techniques, mainly at the expense of providing detailed coverage of the analytic techniques. A notable trend over the past decade, influence and enabled by developments in sensor technology, lower intrusion and size of cameras and recording devices, and, perhaps, attitudes towards and expectations of being recorded in a workplace context, is an increase in application of continuous monitoring or recording of workload in a real workplace context rather than the traditional focus on the laboratory or simulation environment. It is also increasingly common for more than one method of workload assessment to be applied (i.e. triangulation, see Chapter 1).

### ANALYTIC TECHNIQUES

#### Comparability Analysis

This technique is used to prospectively evaluate performance and workload associated with systems that are under development by making comparisons with the performance and workload associated with some reference system. Naturally, it is important that the systems under development have as

many similar characteristics as the reference system as possible. It is difficult to find many references to the use of this technique, but it is often used informally within the design process. However, the use of the technique or tool is encouraged within the MANPRINT framework (Manpower and Personnel Integration – the U.S. Army Human Systems Integration methodology and programme) with regards to supporting the Program Initiation Phase (ODCSPER, 1997). To some extent, the Operational Demand Evaluation Checklist (ODEC) tool applies this approach (Pickup et al., 2010). This tool, developed for the rail signalling context, identifies a number of elements of the rail signalling work that have an influence on the demand placed upon the operator(s). These elements are quantified (e.g. number of stations, number of trains) and on the basis of Subject Matter Expert input have been classified as leading to high, medium or low demand. Therefore, if a new system is being designed, the set of data that now exists to describe current signalling workstations can be compared with the elements to be included in a new workstation design (Pickup et al., 2010).

### Mathematical Models

Information theoretic, manual control and queuing theory models are the most commonly used mathematical models of MWL. They are often not purely used for workload analysis and prediction but have elements or applications that can be used in this context. Information theoretic models are based on the application of information theory (Shannon and Weaver, 1959) to early models of limited processing capacity as described by Welford (1968). The major problem with this approach has been the failure to be able to quantify in information terms for anything more than the most simple of tasks, typified by choice reaction times (Hick, 1952) and simple manual aiming movements (Fitts, 1954). The Keystroke Level Model (Card et al., 1980) combines different elements of interaction, including a mental operator *M*, to predict the time that it will take to complete a task, and derivatives of this are now used in contexts such as automotive control system design. Manual control models are based on classical models of feedback control within engineering systems. They have been reasonably successfully applied to a variety of tracking tasks, both compensatory and pursuit. These tasks form some of the essential components of vehicle driving, aircraft piloting and videogames. Thus, they allow assessments to be made of system controllability within such tasks. However, as already implied, the models are restricted to continuous tasks where the number of input and output variables is limited. Queuing theory models were developed in the context of operations research and have received some success in the context of modelling visual scanning behaviour (Senders and Posner, 1976).

### Expert Opinion

This technique involves taking advantage of experts who have knowledge both of MWL and the work context in which the MWL is to be assessed. A typical example is the use of Pro-SWAT. SWAT stands for Subjective Workload Assessment Technique and, as will be seen in the section on subjective methods, comes in several guises. In the case of Pro-SWAT, the technique parallels the use of the walkthrough technique employed in usability evaluation. An example of the technique is reflected in the study of Kuperman (1985). The context of the study was the conceptual development of a number of alternative fighter aircraft crewstation designs. Several expert pilots were asked to predict workload using the three rating scales typically associated with SWAT (see Table 18.4) in relation to the alternative designs and to a number of typical mission scenarios segments. More frequently, however, expert assessment is sought less formally, and in a more descriptive manner, using tools such as scenarios or Critical Decision Method (CDM) (Klein et al., 1989) to identify elements of a task that may result in a high workload. An example of this was seen in Edwards et al. (2012), who conducted a series of elicitation activities with experts in aviation and ATC to identify factors, including workload, that influenced air traffic controller performance.

### Task Analytic Methods

These methods, probably the most popular of all the analytic methods, initially involve carrying out a task analysis and then generating a timeline analysis on which the workload estimation can

be made. The initial task analysis can be made by using a task specification (task synthesis) or by collecting data from operators performing the task. A common practice is to base the initial task analysis and workload estimation on operators performing the task, and then to propose certain changes to the system design if unacceptably high MWL is identified so that the resulting changes in estimated workload can be predicted. What is important to realise is that the validity of the workload estimation relies on the quality of the task analysis as well as on the workload estimation procedures. Three task analytic models have dominated the literature. They are TLAP (Time-Line Analysis and Prediction; Parks and Boucek, 1989), VACP (Visual, Auditory, Cognitive, Psychomotor; Aldrich et al., 1989) and W/INDEX (Workload Index; North and Riley, 1989). In essence, these three models differ in certain critical underlying assumptions:

- *Channels used:* In completing the timeline analysis, TLAP makes use of five channels: vision, audition (both hearing and speech), hands, feet and cognition. VACP uses the channels implied by the acronym, while W/INDEX uses channels derived from multiple resource theory, typically visual perception, auditory perception, spatial cognition, verbal cognition, manual response and voice response.
- *Red-line threshold:* Both TLAP and VACP models have red-line threshold values above which performance starts to degrade, whereas W/INDEX assumes that performance decrement is a function of task similarity.
- *Demand level:* TLAP assumes demands for any channel are either zero or totally taken up. VACP and W/INDEX permit graded demands.
- *Interaction of workload components/channels:* Both TLAP and VACP assume that the channels represent independent workload components. However, according to the W/INDEX model, resources overlap with each other so that interference can occur across channels as specified by the values found in a so-called conflict matrix.

Over recent years, there have been numerous attempts to validate the various models and to change the various underlying assumptions, particularly in relation to the choice of channels and to the values in the conflict matrix. Peng and ZhiZhong (2014) demonstrated a relationship between VACP analysis and human error rate. Sarno and Wickens (1995) found that both the W/INDEX and VACP models predicted variance in performance in a range of conditions that had different complexities and modalities, but that neither tool captured all elements that might be needed to inform work system design. Examples of domain-specific tools are PUMA (Performance and Usability Modeling in ATM) developed by the National Air Traffic Services for estimating workload associated with ATC activities (Kilner et al., 1998) and ODEC, described previously, and developed for application within rail (Pickup et al., 2010). PUMA relies heavily on the use of video both for the initial task analysis and for representing workload variations over time and incorporates a version of the W/INDEX model.

### Simulation Models

These models attempt to predict human performance by modelling underlying human psychological processes. To a large extent, the validity of such models is a function of how well the models simulate psychological processes. Earlier discussion in this chapter has revealed that, although there is plenty of uncertainty in this area, a number of broad principles have emerged and these form the basis for the simulation models. Probably the most popular simulation modelling tool for ergonomists is Micro Saint (Laughery et al., 2000). It is a general purpose tool, rather than just a simulation of human behaviour, and can be used to model any process that can be represented by a network of tasks or subtasks. More specifically, Micro Saint models the flow of tasks based on specifying factors such as the times to complete tasks, the preconditions under which tasks can be executed, the permitted pathways through the task networks and so on. As with traditional task analysis, the level of detail in the analysis can be decided by the researcher. Typical applications of Micro Saint

to MWL assessment can be found in See and Vidulich (1998) and Liao and Moray (1993); this tool has now been expanded to incorporate elements such as the ACT-R (Adaptive Control of Thought – Rational; Anderson et al., 2004) cognitive architecture to support analysis of dynamic and complex cognitive tasks (Cao and Liu, 2012).

## **EMPIRICAL TECHNIQUES**

### **Primary Task Performance**

When using empirical techniques, it is only natural that measures of primary task performance will be obtained. However, there are a number of obvious limitations with relying solely on these measures. Firstly, while poor human performance can be indicative of task demands being too high, acceptable performance does not necessarily reflect task demands. This is because acceptable performance can be achieved over a range of task demands resulting in varying amounts of spare mental capacity or attentional resources, as illustrated in Figure 18.3. In other words, a person may work harder, but this increased effort may not be apparent from their performance. The amount of available spare capacity is often a critical factor as it determines the available resources to perform other tasks concurrently, particularly those encountered under emergency conditions, that may not be encountered during the MWL estimation procedures. Secondly, if performance measures are obtained over comparatively short periods of time, or periods of time known by the operators, operators can allocate extra resources or engage in task management strategies to cope with periods of exceptionally high task demands. However, because such resources cannot be maintained over extended periods of time, performance will at some point decline. In this respect, MWL is analogous to physical workload where one can only sustain short bouts of very high physical workload before becoming fatigued. Thirdly, and contrary to the previous assertion, there are certain tasks where the allocation of increased resources or effort will not lead to improvements in performance. As mentioned earlier, performance in these tasks is referred to as data limited (Norman and Bobrow, 1975) and are typical of memory retrieval and visual search tasks. Finally, there are problems with the selection and interpretation of performance measures. In some cases, performance measures may be very crude (aircraft lands safely or not safely). In most cases, a range of performance measures is obtained, commonly reflecting performance time and errors. The question is, therefore, how to combine these measures to provide a meaningful performance index. Often indices derived to reflect the speed-error trade-off are based on arbitrary assumptions.

If we consider the role of errors as a primary task measures in the context of the limited resources model, it is clear that they are of most value when an individual is working on a task where their workload capacity is being fully utilised. The value of primary task measures is therefore often in tracking the difference between expertise (e.g. you would expect an expert supermarket checkout operator to be able to process more products in a set time period compared with someone who was unfamiliar with products or technical systems) or the load imposed by different task complexities.

One way to increase understanding of the aforementioned factors is to therefore combine primary task performance measures with others that may capture differences in strategy or perceived effort, such as subjective report tools or expert analysis of strategy and behaviour (see later sections).

### **Secondary Task Techniques**

The development of secondary task techniques, also referred to as dual task or concurrent task techniques, reflects an attempt to quantify MWL by estimating the attentional resources or mental effort invested in primary task performance. It was realised that this could best be done by requiring operators to perform a second task, the secondary task, while at the same time performing the primary task, the task under consideration. The technique aims to quantify, indirectly, spare processing capacity or spare attentional resources (see Figure 18.3). In the original applications of this technique, a comparatively simple view of processing capacity was held, namely that the processing

capacity was limited and that it could be allocated to any task processing component (e.g. perception, response selection, motor control and so on). Two basic forms of the secondary task technique were initially developed. These are the loading task paradigm and the subsidiary (non-loading) task paradigm.

### *Loading Task*

With this paradigm the instructions given to participants emphasise the need to give priority to the secondary task so that performance attains the same level as under a control condition where only the secondary task is given. The workload associated with the primary task is then interpreted in terms of the relative impairment to primary task performance under conditions of concurrent performance compared to when only the primary task is performed. The concurrent secondary task is seen as *loading* the processing capacity and thus leaving less spare capacity available to perform the primary task. Hence, the greater the processing capacity demands associated with the primary task, the greater will be the decrement in primary task performance under dual task conditions.

A typical and still relevant example of the application of this paradigm is illustrated by the results of a study by Dougherty et al. (1964). Under conditions of normal primary flight, performance using a traditional or a newly developed pictorial JANAIR display was indistinguishable. However, adding a secondary task, in this case a simple digit read out task, to which participants had to give priority, was found to impair flight performance much more with the traditional display than with the pictorial display, and this became more evident the greater the difficulty of the loading task. Hence, the addition of the loading task was able to demonstrate that operators had more spare capacity when using the pictorial display, and therefore, that the pictorial display was less demanding than the conventional one. The important implication of this result is that the use of the pictorial display *freed up* processing capacity which could be invaluable under emergency or other high-demand situations.

However, the shortcomings of the technique are very obvious. The technique contravenes the criteria of obtrusiveness because the very aim of the technique is to cause a performance decrement to primary task performance. For this reason, the technique has to be limited to simulator or simulation conditions to avoid the safety consequences of serious errors. Even under these conditions, participants may be reluctant to sacrifice primary task performance by giving priority to a seemingly irrelevant secondary task (hence the importance of obtaining control data from having participants perform both the primary and secondary tasks alone). Second, there is the question of interpreting measures of primary task performance. This has been mentioned in a previous section. However, additional problems of interpretation occur if the conditions being compared do not employ the same primary task performance measures. Schouten et al. (1962), when comparing the workload demands of a number of basic psychomotor tasks, tried to overcome this problem by expressing the decrement in primary task performance under dual task conditions as a percentage of the performance under single-task conditions. However, is it then correct to assume that, say, a 10% increase in the number of errors in the case of one primary task is equivalent to a 10% increase in the time to perform a different type of primary task?

Despite these criticisms, the loading task is a potentially useful technique to assess MWL in the early stages of system design when a realistic prototype is available.

### *Subsidiary (Non-Loading) Task*

The origin of this paradigm is attributed to Brown and Poulton (1961). With this technique, participants are instructed to give priority to the primary task during concurrent performance.\* MWL is

---

\* It is worth noting that the difference between loading and non-loading tasks is principally in the form of instruction given to the participant to concentrate on optimal performance in either the primary or secondary task; this is, therefore, a limitation on this method, in that it depends on the participants adhering to these instructions (and theories of attention tell us that even with a willing and obedient participant, this may be hard to achieve).



then assessed in terms of the impairment to secondary task performance compared to when the secondary task is performed alone. Thus, it is assumed that the more demanding the primary task, the less spare capacity there is available to allocate to the secondary task and, therefore, the greater the related decrement in secondary task performance. The advantages over the loading task paradigm are that this technique allows participants to perform primary tasks under real operational conditions and that MWL can be quantified using the same performance measures that are characteristic of the chosen secondary task. Brown and Poulton (1961) compared the MWL demands of driving a car through a quiet residential area with driving through a busy shopping area. An auditory task was chosen as the secondary task that required the drivers to listen to sequences of eight-digit numbers presented every 4 s. Each sequence of numbers differed from the preceding sequence in respect of one of the eight digits, and it was the task of the driver to identify which number had changed. Not only was it found that the number of errors on this secondary task increased when the task was performed while driving but also that the number of errors was significantly greater when driving through the shopping area compared with the residential area, indicating the relative increased workload. Addressing a similar question, Young and Stanton (1997) applied a visual similarity rotation task as a user-paced, subsidiary task, in a driving simulator study to analyse the impact of automation on the workload of a driving task.

A major difficulty encountered with the application of both these paradigms is that in most cases interference is found between secondary and primary performance which makes quantification of MWL difficult. For example, in the study of Brown and Poulton (1961), under dual-task conditions it was found that slight changes occurred in primary driving performance measures compared with when the driving was performed alone, reflected in reduced driving speeds and changes in control usage. These interference effects are not surprising, and the degree to which they occur is a feature of the extent the primary and secondary task performance involve the same underlying psychological processes. Thus, when comparing the MWL demands of two different primary tasks, one can get different results depending on the particular secondary task selected. It is this feature which gives the secondary task methodology two key characteristics, namely high diagnosticity and high obtrusiveness. Typical secondary tasks include single reaction time, choice reaction time, manual tracking, monitoring, memory, mental arithmetic, speech shadowing and time estimation. Zheng et al. (2010) present a traditional time and error analysis of performance of advanced laparoscopic tasks by surgeons who were experts and novices at a suturing task. They included a secondary visual detection task and found that experts not only performed better on the primary suturing task but also detected more visual signals, and the authors concluded that this evidence suggested that the expert surgeons therefore had spare MWL capacity.

In extreme cases, where the primary and secondary tasks do not share any of the same psychological processes, no interference often occurs so that neither primary nor secondary task performance is degraded under dual-task conditions. Allport et al. (1972) found that skilled pianists could attend to and repeat back continuous speech (speech shadowing) while at the same time sight reading piano music without any degradation in performance to either task compared with the tasks executed alone. Results of this kind lead to the loss of favour of the concept of a single generalised resource of information capacity to be replaced by the development of models based on multiple processing channels and multiple resources (Wickens and Hollands, 1999). These models are also reflected in the development of the task analytic techniques of MWL assessment. This feature is also something that is taken advantage of in the design of some intrusive subjective measures of work – for example, the Situation Present Assessment Method (SPAM) (Durso et al., 1999) measure of situation awareness (see Chapter 19) acknowledges that by its approach of asking questions about current understanding of the current and prospective state of the operating context, it is in itself a form of secondary task – therefore an additional measure that is recorded is the time taken to respond to the question, and there is a protocol for such measures if the participant is so highly loaded on the primary task that they are unable to attend to the subjective questions at all.



Before leaving this category of techniques, two variations should be mentioned that have been developed to overcome some of the problems. These are the adaptive task and the embedded task techniques:

*Adaptive task:* This technique is a variant of the subsidiary technique whereby interference to primary task performance is controlled by varying the difficulty of the secondary task. The cross-adaptive technique developed by Kelly and Wargo (1967) is an example of this. The primary task was a continuous tracking task while the secondary task was a discrete monitoring task. Primary task performance was continuously monitored and, if performance fell below a certain level, the secondary task was turned off until performance on the primary task returned to a reference level. A number of tracking task parameters were varied and the demands placed by these were then assessed in terms of the extent to which the secondary task had to be turned off in order to maintain the same levels of primary task performance. Apart from the technical difficulties of introducing the cross-adaptive technique, the technique is limited to the assessment of primary tasks that are characterised as having a continuous performance component.

*Embedded task:* This is another technique to minimise interference to primary task performance and can be used in operational settings as well as in simulations or laboratory environments. An embedded task is a calibrated task which is already a component of the operator's primary task but which is assumed to take only a secondary role and, therefore, is unlikely to interfere with more important primary task components. Shingledecker et al. (1980) suggested using radio communications activities as the embedded task for fighter pilots. Communications activities chosen for workload estimation required a number of verbal responses and manual radio switching activities to be executed by the pilots. The total time to complete the communications tasks was used as the secondary task performance measure and this was found by Shingledecker and Crabtree (1982) to be sensitive to the loadings imposed upon the primary tracking task performed in a low-fidelity flight simulator. Vidulich and Bortolussi (1988a,b) describe a further application of this technique. Interestingly, the results from their study demonstrated that results based on secondary task performance conflicted with those from using a subjective evaluation technique based on the analytic hierarchy process (AHP) technique (see section 'Relative Judgments'). While the performance data demonstrated benefits from the introduction of speech input devices into a helicopter, the subjective evaluation indicated additional workload imposed by its introduction, possibly from the need to monitor verbal feedback as well as to organise speech inputs.

In general, because of their high obtrusiveness, secondary task techniques are not particularly well suited to the assessment of MWL in realistic operational settings. On the other hand, reflected in their high diagnosticity and sensitivity, the techniques can offer useful insights within laboratory and simulation settings into spare processing capacity and underlying workload components.

## PSYCHOPHYSIOLOGICAL TECHNIQUES

The original assumption underlying the use of psychophysiological measures of MWL is very simple: as workload is increased, there is a corresponding increase in the operator's level of arousal (often referred to as an intervening variable) reflected in the activity of the autonomic nervous system. The level of arousal can be recorded by a number of psychophysiological techniques. While no general relationship between the various psychophysiological variables and known changes in workload has emerged, it has been shown that certain measures demonstrate relative

specificity to different workload components. Because psychophysiological measures are not generally obtrusive, in that participants are not required to execute any extra overt behaviour, and are capable of measuring fluctuations in MWL over time, the techniques offer some distinct advantages over secondary task measures when used as a diagnostic tool. Against this, psychophysiological techniques demonstrate poor selectivity as measures are easily confounded by a variety of external factors including physical workload, emotional stress and environmental factors such as illuminance and noise, previously discussed in terms of the inverted-U relationship between arousal and performance. This means that the techniques are not naturally suited to being used in real operational settings where these variables are hard if not impossible to control for. Despite the enormous reductions in the cost of recording and data analysis hardware and software, the costs are still comparatively high, especially in terms of human support and analysis time. Also, problems associated with low signal-to-noise ratios are frequently encountered. In addition, procedures for measuring and interpreting various psychophysiological measures have not been universally agreed. This is typical of the measurement and interpretation of heart rate variability (sinus arrhythmia). In this book, Parasuraman (Chapter 22) presents an overview of current developments in physiological and neuroergonomics methods; this section briefly describes some of these approaches in the context of workload measurement specifically.

### **Cardiac Activity**

Over the years, cardiac activity has provided some of the most popular psychophysiological measures. To a large extent, this has been because cardiac measures have been relatively easy and cheap to obtain, particularly measures of mean heart rate and heart rate variability. Studies showing the possible sensitivity of mean heart rate to changes in workload include those by Wilson and Fullenkamp (1991) and Bonner and Wilson (2002). In both cases, substantial percentage changes in heart rate above resting levels were associated with different operational segments of a typical flight scenario. In passing, it is worth mentioning that heart rate estimates of MWL did not always correlate with those obtained from subjective evaluations. Additionally, results are not always convincing, to the extent that some studies have failed to demonstrate reliable influences on mean heart rate (e.g. Casali and Wierwille, 1983, 1984; Haapalainen et al., 2010). The lack of consistency in using mean heart rate can be attributed to at least two major factors. Firstly, the data can be seriously confounded by the effects of accompanying physical workload components, both dynamic and static. This is often likely to be the case as there is evidence that the increase in mean heart rate attributable to purely MWL components is very small (Zwaga, 1973). Secondly, and perhaps more significantly, there is a lack of understanding as to why or how mean heart rate (as interpreted as average interbeat interval) should increase with increasing mental demands. One explanation for these inconsistent results is in relation to the so-called intake-rejection hypothesis proposed by Lacey and Lacey (1978). According to this, heart rate increases with the intake of information (e.g. visual detection, scanning or listening) but decreases with the rejection of information (e.g. memory retrieval or problem solving), thus offering some potential diagnosticity to heart rate measures.

Because of the lack of consistency with the results based on mean heart rate (HR), interest has been devoted to measures of heart rate variability (HRV) (sinus arrhythmia) particularly as they are more reflective of the underlying control of heart rate by the autonomic system (both sympathetic and parasympathetic). Measures of heart rate variability range from simple standard deviation of beat to beat intervals, to measures derived from time-series analysis (Billmann, 2011). Mulder (1992) notes that there tended to be a positive correlation between HR and task load for most tasks considered in earlier studies, as opposed to a decrease in HRV with task load. Mulder also highlights the relationship between respiratory rate and HRV, and notes cases where the HRV can potentially be explained by change in respiratory rate; he suggests that it is important to also capture coherence between HRV and respiration.

Brain Activity

Chapter 22 of this book describes the current developments in capture of brain activity in further detail; however, this section summarises application of some example neurological tools to assess workload.

Early work to investigate brain activity and workload focused upon the use of event-related brain potentials (ERPs). The ERP reflects a transient series of oscillations in brain activity that can be recorded via surface electrodes in response to discrete external events. Data analysis requires the relationship between the ERP and the eliciting stimulus to be distinguished from background electroencephalographic (EEG) activity. As with EEG, the ERP has several components but is analysed in the time domain rather than the frequency domain characteristic of EEG analysis. Because the ERP is a response to a specific stimulus, it is often necessary to introduce this stimulus as a separate secondary task, although it is possible to analyse the ERPs related to discrete stimulus events embedded within the primary task. Hankins and Wilson (1998) found that the theta band of the EEG (see Chapter 21 for further explanation of the interpretation of EEG frequency bands) was particularly effective at highlighting elements of an aviation task which required inflight mental calculations.

More recent work has focused on the value of detailed measures that capture the different areas of the brain involved in cognitive activity, such as functional magnetic resonance imaging (fMRI) and functional near infrared spectroscopy (fNIRS). Pike et al. (2014) summarise the practical features of three different brain sensing techniques – fMRI, EEG and fNIRS (Table 18.2).

This highlights a particular feature that is of importance for E/HF studies – sensitivity to motion. Most E/HF studies have a level of complexity that is sufficient that it is not appropriate for the participant to be held physically still during task completion; indeed, in some contexts, there is considerable additional information to be gleaned from the physical motion of the participant (see later discussion of gait analysis for example).

It is likely that as technical advancements lead to continued reduction in the cost and intrusivity of techniques such as fNIRS they will become more feasible for implementation in the workplace context; therefore of considerable importance is a clear understanding of the methods of interpretation of data from such tools, and their sensitivity and diagnosticity to changes in experiences of workload. Peck et al. (2013) found a negative correlation between de-oxygenated haemoglobin (Hb) from fNIRS and subjective rating of mental demand; whereas Pike et al. (2014) did not find a similar relationship, but did identify a positive correlation between Hb and performance. This perhaps suggests that when mental demands are high (and probably beyond the point where limited resources capacity is reached), performance and Hb decrease.

**TABLE 18.2**  
**Comparison Summary of Three Example Brain-Sensing Techniques for Workload Assessment**

| Technique                  | Physical Property | Sensitivity to Movement/Motion | Portability of Equipment | Spatial Resolution | Temporal Resolution |
|----------------------------|-------------------|--------------------------------|--------------------------|--------------------|---------------------|
| fMRI                       | Magnetic          | Very high                      | None                     | High               | Low                 |
| Electroencephalogram (EEG) | Electrical        | High                           | Portable                 | Low                | High                |
| fNIRS                      | Optical           | Low                            | Portable                 | High               | Low                 |

*Source:* Adapted from Pike, M. et al., Measuring the effect of think aloud protocols on workload using fNIRS, in *Proceedings of CHI 2014 ACM*, Toronto, Ontario, Canada, 26 April–1 May 2014.

### Electrodermal Activity

Electrodermal activity, and galvanic skin response (GSR) in particular, has been applied to measure workload. Reimer and Mehler (2011) demonstrated that skin conductance and heart rate both increased with increased cognitive demand in a study that captured physiological response in both a field and simulator context. Haapalainen et al. (2010) found that heat flux (rate of heat transfer measured using an instrumented armband) was a more effective measure of cognitive demand than GSR, although other authors (e.g. Kilseop and Rohae, 2005; Shi et al., 2007) have suggested that GSR is a useful indicator of workload.

### Eye Function

Improvements in quality of equipment to capture and tools to analyse eye movements and responses have led to an increase in the use of eye function measures as indicators of workload. Such measures include endogenous eye blinks (those which are not made reflexively to specific stimuli) and usually capture blink frequency and blink duration.

Blink rate has been found to decrease under conditions of high workload both in driving tasks (Lecret and Pottier, 1971) and in flight tasks (Wilson and Fullenkamp, 1991; Wilson et al., 1987). However, the relationship is not always as expected so that in some cases blink rate is higher under relatively high workload conditions, possibly because there is a tendency to blink after taking in visual information. Another study failed to reveal any effects on blink rate (Casali and Wierwille, 1983). In their study, auditory rather than visual load was varied. Merat et al. (2012) and Recarte et al. (2008) propose that for a task that has a higher visual demand, blink rate and duration are decreased as participants attempt to capture as much visual information as possible (i.e. demonstrating blink inhibition); however, when a non-visual but cognitively demanding task is present, this blink inhibition is suppressed, and in fact an increase in blink rate and/or duration may be seen.

Over the centuries, market traders have assessed a person's interest in an object they are trying to sell by observing his or her pupil diameter. With increased interest and arousal, pupil size increases. The use of pupil size as a measure of MWL goes back to Kahneman (1973), who demonstrated that it was closely associated with the number of items held in short-term or working memory. Since then, a number of studies have shown pupil size to be sensitive to a number of workload components including perceptual (Beatty, 1982a,b), cognitive (Casali and Wierwille, 1983) and response-related (Richer and Beatty, 1985) components. Schulz et al. (2011) found an increase in pupil diameter with increasing severity of simulated clinical incident. Thus, while pupil size appears to meet the sensitivity criterion, it demonstrates little diagnosticity. It is claimed, however, that it can distinguish data-limited processing from resource-limited processing (Beatty, 1982a,b).

The main problems with pupil size measurements are that they are confounded by ambient illuminance and by the combination of illuminances of various light sources within the field of view. Additionally, pupil size is influenced by eye accommodation and vergence processes associated with depth perception. For these reasons, the use of pupil diameter as a measure of workload in real operational settings is likely to remain limited.

### Body Fluid Analysis

Analysis of hormonal or blood glucose levels has also been suggested as indicators of MWL, but their use is less frequent, normally due to the complexities in sample collection and delay between sample collection and analysis point (i.e. only a few point-based measures therefore being possible, as opposed to continuous or more frequent measures which are possible from some other physiological measures). For example, Shelton-Rayner et al. (2012) demonstrated that changes in leukocyte activity were associated with subjective reports of mental load. Engelmann et al. (2011) found that testosterone and cortisol levels were lower in surgeons who had been given more small breaks during a task, with an aim to reduce mental fatigue.

### Muscle and Movement Analysis

The final class of psychophysiological measures to be considered are those that capture movement or physical effort. The relationship between physical and mental load has been subject to much discussion and debate. Mehta and Agnew (2013) suggest that increased MWL was associated with a reduction in agonist EMG activity. In addition to measuring muscle activity, there is increasing interest in the relationship between posture and gait and workload – if a strong relationship can be demonstrated, this provides a basis to interpret data captured through observation or measurement as indicative of subjective experience of workload, without the need for application of intrusive question-based measures in a workplace context. The relationship between captured posture and activity is highlighted by Sharples et al. (2011, 2012), where a distinction between active monitoring (leaning forwards at the workstation) and passive monitoring (in a reclining position) was found to have some association with subjective reports of workload. Witchel et al. (2013) suggest that measured changes in head height and shoulder height may indicate loss of vigilance or diminishing arousal in seated computer users. This is an area that new sensing technologies and computer vision techniques will allow more detailed consideration of in the near future.

### SUBJECTIVE/OPERATOR OPINION TECHNIQUES

Finally, we come to the ergonomist's favourite class of techniques! Subjective measures of MWL keep growing in popularity. Not only are they easy and cheap to administer, but they are also characterised by high face validity – when somebody tells you that they feel they have had to put a lot of effort into performing a task, one cannot help but assume that this experience reflects the execution of processes underlying task performance. In this context, the participants are judging the interaction between themselves and the task demands. On the other hand, true validity remains relatively elusive. The general use of subjective techniques in E/HF, including rating scales, is covered in Chapter 4, while Eggemeier and Wilson (1991) present an excellent review of the development of subjective measures of MWL.

Three rating scales have dominated the literature. The first of these is the Cooper–Harper Scale (Cooper and Harper, 1969) which employs a single 10-point scale, each point having a verbal descriptor, while the other two are multidimensional scales, NASA Task Load Index (TLX) (Hart and Staveland, 1988) and SWAT (Reid and Nygren, 1988). A number of modifications have been made to the Cooper–Harper Scale, the two most frequently used versions being the Modified Cooper–Harper Scale (Wierwille and Casali, 1983) and the Bedford Scale (Roscoe, 1987). Numerous other scales have been developed, incorporating various degrees of methodological rigour, and some of these will be described in addition to those already mentioned. The different subjective techniques can be distinguished by a number of features: (1) single or multiple dimension scales, (2) absolute or relative judgements and (3) retrospective or instantaneous data collection. In general, Gopher and Braune (1984) have provided good evidence of the potential high consistency of subjective estimates of workload.

#### Single-Dimensional Scales

The *Cooper–Harper Scale* (*Cooper–Harper Aircraft Handling Characteristics Scale*; Cooper and Harper, 1969) is still widely used within the aircraft industry, for which it was designed. It is especially tailored towards performance and controllability in relation to aircraft characteristics. The 10-point scale is accompanied by verbal descriptors for each of the points. For example, a pilot rating of 6 is equivalent to the verbal descriptor *adequate performance requires extensive pilot compensation*. This is taken to indicate that the aircraft has very objectionable but tolerable deficiencies and that these warrant improvement. An obvious feature of the scale is that no references are made in the verbal descriptors to workload.

The *Modified Cooper–Harper Scale* (Wierwille and Casali, 1983) consists of a scale where descriptors have been reworded in terms of the mental effort required to perform tasks and,



therefore, can be applied to tasks other than controlling an aircraft. For example, the verbal descriptor for point 6 of the scale now reads *maximum operator mental effort is required to attain adequate system performance*, indicating that MWL is high and should be reduced.

The *Bedford Rating Scale* (Roscoe, 1987) was also developed for application within the flight environment. Again, the 10 original workload level descriptors from the Cooper–Harper Scale have been modified. Interestingly, a majority of the descriptors make reference to spare capacity rather than or as well as effort invested. Referring to point 6 of the scale, the verbal descriptor reads *little spare capacity: level of effort allows little attention to additional tasks*, indicating that workload needs to be reduced. Although descriptors of this kind match up with some definitions of MWL, it is quite difficult to see how participants can estimate their own spare capacity, especially as spare capacity seems likely to be a function of the specific processing resources involved. Nevertheless, numerous studies using the scale have confirmed that it can be used to differentiate between different levels of loading in simulated flight environments (e.g. Tsang and Johnson, 1989).

### Multidimensional Scales

Because most contemporary definitions of MWL refer to its multidimensional characteristic, it is not surprising that attempts have been made to capture this aspect by developing multidimensional assessment procedures. As well as the possibility of developing more valid assessments of workload, such procedures have the additional potential of providing a diagnostic tool.

The *NASA-TLX* (Hart and Staveland, 1988) is the most widely used of all MWL scales, not only because of its multidimensional feature but also because it is comparatively easy to administer. The technique requires participants to complete ratings in relation to six different scales, each scale represented by a 10 cm line divided into 20 intervals. The ratings on each scale are converted to values of 0–100. The scales were chosen by Hart and Staveland (1988) as a result of an extensive programme of laboratory research. The definitions of the scales are given in Table 18.3. A notable feature of the scales is that a physical workload dimension is included. Following the ratings, which are made retrospectively, participants are required to weight the relevance of the six dimensions to the task under investigation by completing a paired comparison procedure. All pairs of the dimensions are compared, 15 in all, so that a particular dimension could end up with a weighting of a value between 0 and 5. The overall workload index is computed by combining the ratings and

**TABLE 18.3**  
**The NASA-TLX Rating Scale Definitions**

| Dimension       | Endpoints | Descriptors                                                                                                                                                                                                          |
|-----------------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mental          | Low/high  | How much mental and perceptual activity was required DEMAND (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving? |
| Physical        | Low/high  | How much physical activity was required (e.g. pushing, DEMAND pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous?                                   |
| Temporal demand | Low/high  | How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?                                                      |
| Performance     | Good/poor | How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?                        |
| Effort          | Low/high  | How hard did you have to work (mentally and physically) to accomplish your level of performance?                                                                                                                     |
| Frustration     | Low/high  | How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?                                                                   |



**TABLE 18.4**  
**SWAT Rating Scale Definitions**

1. Time load
  - a. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
  - b. Occasionally have spare time. Interruptions or overlap among activities occur frequently.
  - c. Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.
2. Mental effort load
  - a. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.
  - b. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability or unfamiliarity. Considerable attention required.
  - c. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.
3. Psychological stress load
  - a. Little confusion, risk, frustration or anxiety exists and can be easily accommodated.
  - b. Moderate stress due to confusion, risk, frustration or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
  - c. High or very intense stress due to confusion, frustration or anxiety. High to extreme determination and self-control required.

the weights: each of the 6 ratings is multiplied by the corresponding weight to give adjusted ratings, these are then summed and, finally, this sum is divided by 15 to give a value between 0 and 100. The need to include the weighting procedure, which naturally prolongs the data collection time, has been queried. Results from Byers et al. (1989) and Dickinson et al. (1993) suggest the weighting protocol is unnecessary as very high correlations were found between weighted and unweighted scores, although, as Dickinson et al. (1993) also point out, the correlations were not perfect, suggesting the weighting protocol may make a small but significant contribution.

Compared with the NASA-TLX, the SWAT (Reid and Nygren, 1988) is time consuming, often requiring an hour or so to implement fully. A near unique feature of SWAT is that it is based on a psychological model of how judgements of MWL are formed by participants. The technique requires participants to rate on three workload dimensions, each rating scale having three points. These are described in Table 18.4. However, before this is done, participants have to develop an underlying scale of MWL. This is achieved by asking participants to sequence 27 cards in relation to the workload level they represent. Each card is given a unique combination of the three levels of each of the three dimensions shown in Table 18.4, with 1-1-1 naturally being allocated the lowest position in the sequence (lowest workload) and 3-3-3 the highest position (highest workload). While performing this scaling procedure to order the remaining 25 cards, participants have to imagine a task or work situation that they feel corresponds to the various combinations of levels of the three dimensions. The sequence of cards is then subjected to a conjoint measurement technique developed by Krantz and Tversky (1971), which produces an interval scale of workload ranging from 0 to 100. This means that each of the 27 combinations of levels of the three workload dimensions now has a value of between 0 and 100 associated with it. Once this has been done, all that remains is for the participant to rate the task under examination in relation to the three dimensions so that the corresponding workload value can be read off from the previously developed underlying workload scale.

The need to go through the conjoint measurement procedure for each participant results from the finding that participants do not give equal weighting to the three dimensions during this procedure (e.g. some participants might weigh time load as being the more important dimension while others rate effort or stress as being the more important) (Reid and Nygren, 1988). Hence, the weighting procedure, like that used by NASA-TLX, can be seen as a way of reducing individual differences. However, Biers (1995) has claimed that simply using the composite scores from the three SWAT scales and omitting the weighting procedure (i.e. so that workload is effectively rated on a 7-point

scale with unitary values going from 3 to 9) does not lead to a decrease in sensitivity. Biers' study involved fighter pilots and manipulated cockpit configuration and mission task. In its favour, the use of the conjoint measurement procedure may enable the SWAT technique to be used in a predictive manner, as in the case of Pro-SWAT.

Tsang and Velazquez (1996) propose the *Workload Profile (WP)* – a tool that is specifically derived from Wickens' multiple resources model, and asks participants to rate, on a scale between 0 and 1, the different demand that they are experiencing along the dimensions of stage of processing (perceptual/central), code of processing (spatial/verbal), input (visual/auditory) and output (manual/speech). Although this does rely on the participant's interpretation of the descriptors of the different parameters, its strength lies in its direct relevance to the underlying theoretical model and the acknowledgement of the different cognitive functions. This distinction between functions is of course different from the different dimensions in NASA-TLX or SWAT, which are more closely related to the way in which a task is implemented in a workplace context (e.g. time pressure).

### Relative Judgements

All the scaling techniques described so far in this section have involved absolute judgements, which means there is considerable uncertainty concerning the characteristics of the underlying scales. In most cases, it would seem safest to conclude that these scales are ordinal, although protagonists of SWAT have argued that they reflect interval scales. The use of relative judgement techniques is one way to be more confident that the underlying scales are indeed interval ones. This can be achieved by using paired comparison and magnitude estimation techniques. An example of the use of paired comparisons is the AHP developed by Saaty (1980) and implemented for workload assessment by Vidulich and Tsang (1987). As Vidulich (1989) points out, AHP does not '... use raters as workload meters'. On the other hand, AHP is used to elicit '... the conscious decision-making and experiential knowledge to extract expert judgements about workload'. The AHP forms the basis for the subjective technique developed by Vidulich and his colleagues (Vidulich 1989; Vidulich et al., 1991) entitled SWORD (Subjective WORKload Dominance). The SWORD technique involves three stages: (1) collecting the paired comparison data, (2) constructing a judgement matrix and (3) calculating the SWORD workload ratings. The paired comparisons technique involves raters comparing all pairs of tasks or design alternatives on a 17-point scale where point 9 represents equal rating of the two alternatives. From the subsequent judgement matrix, workload ratings are calculated by obtaining the geometric mean for each row of the matrix and then normalising the means. A consistency value is also obtained which assesses the extent to which the underlying workload scale represents an interval scale. Vidulich and his colleagues were particularly interested in using SWORD as a projective or predictive technique. One of the examples used by Vidulich et al. (1991) was the evaluation of a variety of proposed formats for aircraft HUDs. Results showed that the ratings of predictive MWL by pilots correlated very well with retrospective ratings and that both types of rating correlated well with performance under simulation conditions. The authors argue that the SWORD technique may be a viable alternative to task analytic techniques, but with the added advantage of being considerably less labour intensive.

### Instantaneous Judgements

All the scaling techniques mentioned so far have involved retrospective or prospective judgements. While this feature does not appear to have invalidated their use, many investigators have wanted to be able to see how workload varies during the course of performing a task without resorting to psychophysiological techniques. For this to be achieved, the ratings need to be completed at frequent intervals and as quickly as possible while the primary task is being performed. In a sense, the rating procedure can now be seen as a concurrent secondary task.

Pilot Objective/Subjective Workload Assessment Technique (POSWAT) is an example of this technique where pilots have to make judgements every minute using a kneeboard with 10 keys, key 1 corresponding to little or no workload and key 10 to excessive workload where the pilot can

only just maintain flying the aircraft. The pilot is prompted to make his or her judgements by a tone presented over the pilot's headset and a red light appearing on the kneeboard. Unfortunately, Mallery (1987) does not present the results on a minute-to-minute basis, but rather he averages ratings for each of the flight scenario segments. These average scores were found to be sensitive to the different demands associated with the segments.

Instantaneous Self-Assessment (ISA) is a further example of this technique and was developed by the National Air Traffic Services (UK) and Eurocontrol to provide continuous ratings of MWL during ATC tasks. The scale is a 5-point one and uses a small keypad with the five keys corresponding to *very high*, *high*, *fair*, *low*, *very low* (previously *excessive*, *high*, *comfortable*, *relaxed* and *under-utilised*). Participants are cued to make a rating by a visual signal presented on the radar console.

A major concern with these techniques, particularly if they are to be used under real operational conditions, is that the execution of the ratings may act as an extra concurrent task, not only influencing primary task performance but also leading to an overall increase in workload ratings. To see if primary task performance is influenced by concurrent workload judgements, Tattersall and Foord (1996) investigated the use of ISA in a laboratory setting. Participants made judgements every 2 min while performing a simple pursuit tracking task with three levels of task difficulty. Half the participants used a keypad to make their judgements while the other half spoke their responses. In addition to using the ISA technique, retrospective subjective ratings were taken using a modified SWAT procedure where the weighting protocol was excluded. Essentially the results from the ISA technique correlated very closely with the ratings on the three SWAT scales. However, most importantly, the results showed that the tracking error was increased at the time the judgements were made. Contrary to the predictions of multiple resource theory, verbal responses caused the same decrement in the primary tracking performance as the manual responses, suggesting the added demand created by the rating procedure was not simply one of competing output modes, in which case one should have observed more interference when using the keyboard to input the workload ratings.

Finally, the IWS has been developed for application in a rail context (Pickup et al., 2005). The IWS was designed to capture workload in rail signallers and also aimed to be in a simple form that could be regularly responded to in a real and simulated workplace context. After initial review of other single-dimensional scales, it became apparent that the words of some of the scales noted earlier were not appropriate for the rail context – in particular, rail signallers commented that, due to the physical interlocking and safety systems present in the rail context, it was not possible for them to be operating the system in a state that was *unsafe*,\* and they felt uncomfortable with the words *unsafe* within the scale. Therefore, a new scale was developed using a Thurstone rating technique (see Chapter 4 for description of the Thurstone approach applied to develop the IWS).

### Comparison between Rating Scales

Numerous studies have compared the use of the different scales, mainly in relation to sensitivity. At this stage, one can do little more than echo the remarks of Eggemeier and Wilson (1991) when they conclude that there is insufficient evidence to favour any one particular scale. However, there are the obvious benefits of employing multiple dimensional ratings in that they have some diagnostic value as well as aiding predictive workload evaluation. There are also the claims made that if the weighting procedures are employed, both SWAT and NASA-TLX demonstrate less individual variations in ratings than are found with the single-dimensional scales. When it comes to participants' acceptance, there is some evidence from the results of Byers et al. (1988) that when the weighting procedures are omitted from both NASA-TLX and SWAT, NASA-TLX is more acceptable than both SWAT and the Modified Cooper–Harper Scale. In an early comparison study by Wewerinke (1974), reported by Moray (1982), three popular single-dimensional scales were compared: a simple non-verbal 10 cm line requiring an estimate of *effort you spend performing the task* from a value of

---

\* Although there are occasions in U.K. rail where the signaller is required to give instructions that override the physical interlocking systems present (e.g. in cases of technical failure).

0–10, the original Cooper–Harper Scale and the modified Cooper–Harper Scale. The correlations between the ratings from the three scales were nearly perfect, suggesting the verbal descriptors used in the Cooper–Harper Scales were adding little or nothing to the assessment process.

Rubio et al. (2004) evaluated the intrusiveness, sensitivity, diagnosticity and validity of the NASA-TLX, SWAT and a third measure – the WP. They found that all three measures were similar in terms of intrusiveness, had high concurrent validity and that, as would be expected from its theoretical derivations, the WP was particularly effective in terms of diagnosticity.

## Interviews

An oft-overlooked method of assessing workload is the use of interviews (see Chapter 4). These are of course subjective both from the point of view that the information is elicited from the participant, but also potentially introduce subjectivity on the part of the E/HF researcher. Interviews that aim to capture aspects of workload can use approaches such as scenarios or use cases, structured elicitation techniques such as CDM (Klein et al., 1989) or simply be exploratory and semi-structured. They are often most useful to inform the design of specific conditions to be implemented in a simulated context, to provide insight into the appropriate times of day or circumstances in which it might be most fruitful to apply other methods such as physiological assessment or primary task analysis or to explain causative factors behind workload phenomena that have been captured or hypothesised. The method of reporting such data is inevitably qualitative and descriptive, but the use of specific quotes or content analytic techniques (e.g. theme-based content analysis; Neale and Nichols, 2000) can be useful to reinforce data from other, quantitative, sources.

One particular area in which interviews have been applied to elicit indicators of workload is in work that has aimed to capture behavioural markers. Such markers are indicative of different workloads and have been identified in the surgical (Fletcher et al., 2004) and ATC contexts (Edwards et al., 2012) and have potential to be very useful in supporting supervisory tasks and identifying the types of behaviours that an individual or team might exhibit when experiencing high or low workload.

## Observations

A final subjective method to consider is that of observation. In addition to being used as part of the methodological suite that inform behaviour markers (e.g. Fletcher et al., 2004), they can be used to support inferences about the demand placed upon individuals by different work system designs or automation levels. Sharples et al. (2011) describe a study where a set of six main categories of observable activity were captured through expert observation and demonstrated that there was some relationship between observed interaction and task design; in this form, observation acts as an empirical form of the task analytic methods described earlier in this chapter. Sharples et al. (2012) note, however, that such tools may depend on the *observability* of an individual operator; some operators may be more overt in the extent to which they demonstrate their experienced workload through observable interaction compared to others.

## Validity of Subjective Measures

It has generally been agreed that the validity of subjective measures should be discussed mainly in terms of sensitivity to variations in task demands and but also in relation to primary task performance. There are limitations with these criteria. Firstly, sensitivity is usually assessed in relation to the investigators *expert* evaluation of task difficulty or task loading. However, definitions of MWL frequently express workload not in relation to task loading per se but to loading of the individual, often in relation to some underlying model of information processing. Similarly, it has been argued that primary task performance may not necessarily reflect workload to the extent that workload and subjective ratings may become dissociated. That is to say, it does not follow that as levels of subjective workload increase, primary task performance will become correspondingly impaired. A number of these dissociations have been described by Yeh and Wickens (1988). In principle,

these dissociations are derived from the reasonable assumption that subjective ratings are reflecting the amount of attentional resources or effort a person is investing in performing a particular task:

1. In cases of task underload, where available resources far outweigh the resources required by the task demands, individuals may invest unnecessarily greater resources, thinking this will improve their performance. An example of this kind of argument is supported by the results when using the NASA-TLX to assess the MWL associated with vigilance tasks. These tasks are traditionally classified as undemanding, yet the results of Warm et al. (1996) confirm that surprisingly high workload ratings are reported especially in relation to the mental effort and frustration scales. The authors suggest that extra resources above those required by the task demands are invested to overcome the tedium accompanying such tasks. An alternative interpretation of this phenomenon is that performance is, in effect, data- rather than resource-limited (Norman and Bobrow, 1975). In the case of overload, where resources are fully invested, any increase in task demands will be reflected in decreased performance but no increase in subjective assessment because resources are fully invested. In other words, the subjective measures have saturated.
2. Because greater investment of resources should yield better performance for resource-limited tasks, there is an inherent dissociation whereby increased resource investment leads to superior performance. This can be achieved by increasing an individual's motivation, for example by providing performance feedback.
3. Dissociation will occur when an easy dual task configuration is compared to a difficult single task. The example cited by Yeh and Wickens (1988) is one of introducing automation to aid primary task performance. Although improvement in performance is likely to be found, subjective rating may be higher for the aided situation because of the need to invest increased resources as a result of the time-sharing aspect of the aided task, most probably from an increase in working memory load; that is to say, an increase in task management demands. This is confirmed by Aretz et al. (1996). Using the NASA-TLX, they demonstrated that out of a number of potential task demands, the number of concurrent tasks had the largest impact on pilots' subjective workload ratings.
4. According to multiple resource theory (Wickens and Hollands, 1999), performance in multiple task environments is likely to be worse when the tasks compete for similar resources. This rivalry for resources will not be reflected in many of these subjective measures because the overall amount of resources invested may be the same as when there is not such high competition for resources.

A strong argument, therefore, can be put forward for placing limited reliance on the exclusive use of subjective ratings of MWL in multitask conditions and on using such ratings in a predictive manner.

## CONCLUSIONS

Readers can be forgiven for concluding that MWL evaluation presents a potential minefield and that it is necessary to proceed with great caution. What we have aimed to do in this chapter is provide a very general framework (Figure 18.1) for the evaluation process in the absence of a universally agreed definition of MWL and in a climate of continuing research into underlying psychological processes. We have overviewed a majority of the measurement techniques in relation to the framework and have indicated their relative advantages and disadvantages according to a number of well-established criteria. What can be concluded with confidence is that the principle of triangulation is particularly important when capturing workload in either simulated, laboratory or real-world settings; in order to gain insight into MWL issues, there is a need to use a range of measurement techniques rather than *putting all your eggs into one basket*.



## REFERENCES

- Aldrich, T.B., Szabo, S.M. and Bierbaum, C.R. (1989). The development and application of models to predict operator workload during system design. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton and L. van Breda (Eds.), *Applications of Human Performance Models to System Design*, pp. 65–80. New York: Plenum Press.
- Allport, D.A., Antonis, B. and Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, **24**, 225–235.
- Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C. and Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, **111**(4), 1036–1060.
- Aretz, A.A., Johannsen, C. and Ober, K. (1996). An empirical validation of subjective workload ratings. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, pp. 91–95. Santa Monica, CA: The Human Factors and Ergonomics Society.
- Balfe, N. (2010). Appropriate automation of rail signalling systems: A human factors study. PhD thesis, University of Nottingham, Nottingham, U.K.
- Beatty, J. (1982a). Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology*, **19**, 167–172.
- Beatty, J. (1982b). Task-evoked pupillary responses, processing loads, and the structure of processing resources. *Psychological Bulletin*, **91**, 276–292.
- Biers, D.W. (1995). SWAT: Do we need conjoint measurement. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, pp.1233–1237. Santa Monica, CA: The Human Factors and Ergonomics Society.
- Billman, G.E. (2011). Heart rate variability – A historical perspective. *Frontiers in Physiology*, **2**, 86.
- Bonner, M.A. and Wilson, G.F. (2002). Heart rate measures of flight test and evaluation. *International Journal of Aviation Psychology*, **12**, 63–77.
- Brown, I.D. and Poulton, E.C. (1961). Measuring the spare ‘mental capacity’ of car drivers by a subsidiary task. *Ergonomics*, **4**, 35–40.
- Byers, J.C., Bittner, A.C., Hill, S.G., Zaclad, A.L. and Christ, R.E. (1988). Workload assessment of a remotely piloted vehicle (RPV) system. In *Proceedings of the Human Factors 32nd Annual Meeting*, pp. 1145–1149. Santa Monica, CA: The Human Factors Society.
- Cao, S. and Liu, Y. (2012). QN-ACTR modeling of multitask performance of dynamic and complex cognitive tasks. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*, pp. 1015–1019. Santa Monica, CA: The Human Factors and Ergonomics Society.
- Card, S.K., Moran, T.P. and Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, **23**(7), 396–410.
- Casali, J.G. and Wierwille, W.W. (1983). A comparison of rating scale, secondary task, physiological, and primary task workload estimation techniques in a simulated flight emphasising communications load. *Human Factors*, **25**, 623–641.
- Casali, J.G. and Wierwille, W.W. (1984). On the measurement of pilot perceptual workload: A comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics*, **27**, 1030–1050.
- Chandler, P. and Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, **8**(4), 293–332.
- Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, **25**(5), 975–979.
- Cooper, G.E and Harper, R.P. (1969). The use of pilot rating in the evaluation of aircraft handling qualities. Report No. NASA TN-D-5153. Ames Research Centre, Moffett Field, CA.
- Cox, T. (1978). *Stress*. London, U.K.: MacMillan Press.
- Cox, T. and Mackay, C.J. (1981). A transactional approach to occupational stress. In E.N. Corlett and J. Richardson (Eds.), *Stress, Work Design and Productivity*. Chichester, U.K.: Wiley & Sons.
- Dickinson, J., Byblow, W.D. and Ryan, L.A. (1993). Order effects and the weighting process in workload assessment. *Applied Ergonomics*, **24**, 357–361.
- Dougherty, I., Emery, J.H. and Curtin, J.G. (1964). Comparisons of perceptual work load in flying standard instrumentation and the contact analog vertical display (JANAI-R-D228-421-019), Bell Helicopter Co., DTIC No. AD610617.
- Durso, F.T., Hackworth, C.A., Truitt, T.R., Crutchfield, J. and Nikolic, D. (1999). Situation awareness as a predictor of performance in en route air traffic controllers. No. DOT/FAA/AM-99/3. Department of Psychology, Oklahoma University, Norman, OK.



- Edwards, T., Sharples, S., Wilson, J.R. and Kirwan, B. (2012). The need for a multi-factorial approach to safe human performance in air traffic control. In *Proceedings of the Fourth AHFE International Conference*. San Francisco, CA, 21–25 July 2012.
- Eggemeier, F.T. and Wilson, G.F. (1991). Performance-based and subjective assessment of workload in multi-task environments. In D.L. Damos (Ed.), *Multiple-Task Performance*, pp. 217–278. London, U.K.: Taylor & Francis Group.
- Engelmann, C., Schneider, M., Kirschbaum, C., Grote, G., Dingemann, J., Schoof, S. and Ure, B.M. (2011). Effects of intraoperative breaks on mental and somatic operator fatigue: A randomized clinical trial. *Surgical Endoscopy*, **25**(4), 1245–1250.
- Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, **47**, 381–391.
- Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N. and Patey, R. (2004). Rating non-technical skills: Developing a behavioural marker system for use in anaesthesia. *Cognition, Technology & Work*, **6**(3), 165–171.
- Gopher, D. and Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, **26**, 519–532.
- Haapalainen, E., Kim, S.-J., Forlizzi, J.F. and Dey, A.K. (2010). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, pp. 301–310. New York: ACM.
- Hankins, T.C. and Wilson, G.F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, **69**(4), 360–367.
- Hart, S.G. and Staveland, L.E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*, pp. 139–183. Amsterdam, the Netherlands: North-Holland.
- Hick, W.E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, **4**, 11–26.
- Hill, S.G. et al. (1987). Analytic techniques for the assessment of operator workload. In *Proceedings of the 31st Annual Meeting of the Human Factors Society*, pp. 368–372. Santa Monica, CA: The Human Factors Society.
- International Organization for Standardization (2000). ISO 10075-1, Ergonomic principles related to mental work-load, part 1: General terms and definitions. Geneva, Switzerland: ISO.
- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kelly, C.R. and Wargo, M.J. (1967). Cross-adaptive operator loading tasks. *Human Factors*, **9**, 395–404.
- Kilner, A., Hook, M., Fearnside, P. and Nicholson, P. (1998). Developing a predictive model of controller workload in air traffic management. In M.A. Hansen (Ed.), *Contemporary Ergonomics 1998*, pp. 409–413. London, U.K.: Taylor & Francis Group.
- Kilseop, R. and Rohae, M. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, **35**(11), 991–1009.
- Klein, G.A., Calderwood, R. and Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, **19**(3), 462–472.
- Krantz, D.H. and Tversky, A. (1971). Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, **78**, 151–169.
- Kuperman, G.G. (1985). Pro-SWAT applied to advanced helicopter crewstation concepts. In *Proceedings of the 29th Annual Meeting of Human Factors Society*, pp. 398–402. Santa Monica, CA: The Human Factors Society.
- Lacey, J.I. and Lacey, B.C. (1978). Two-way communication between the heart and the brain: Significance of time within the cardiac cycle. In E. Meyer and J. Brady (Eds.), *Research in the Psychobiology of Human Behavior*, pp. 99–113. Baltimore, MD: Johns Hopkins University Press.
- Laughery, R., Archer, S., Plott, B. and Dahn, D. (2000). Task network modelling and the Micro Saint family of tools. In *Proceedings of the 44th Annual Meeting of Human Factors and Ergonomics Society*, pp. 721–724. Santa Monica, CA: The Human Factors and Ergonomics Society.
- Leclerc, F. and Pottier, M. (1971). La vigilance, facteur de securite dans la conduite automobile. *Le Travail Humain*, **34**, 51–68.
- Liao, J. and Moray, N. (1993). A simulation study of human performance deterioration and mental workload. *Le Travail Humain*, **56**, 321–344.
- Linton, P.M. et al. (1989). Operator workload for military system acquisition. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton and L. van Breda (Eds.), *Applications of Human Performance Models to System Design*, pp. 21–45. New York: Plenum Press.

- Mallery, C.J. (1987). The effect of experience on subjective ratings for aircraft and simulator workload during IFR flight. In *Proceedings of the Human Factors Society 31st Annual Meeting*, pp. 838–841. Santa Monica, CA: The Human Factors Society.
- Mayer-Schonberger, V. and Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London, U.K.: John Murray Publishers.
- Mehta, R.K. and Agnew, M.J. (2013). Exertion-dependent effects of physical and mental workload on physiological outcomes and task performance. *IIE Transactions on Occupational Ergonomics and Human Factors*, **1**(1), 3–15.
- Merat, N., Jamson, A.H., Lai, F.C.H. and Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **54**, 762.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, **7**(3), 134–140.
- Moray, N. (ed.) (1979). *Mental Workload: Its Theory and Measurement*. New York: Plenum Press.
- Moray, N. (1982). Subjective mental workload. *Human Factors*, **24**, 25–40.
- Mulder, L.J.M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, **34**, 205–236.
- Neale, H. and Nichols, S. (2001). Theme-based content analysis: A flexible method for virtual environment evaluation. *International Journal of Human-Computer Studies*, **55**(2), 167–189.
- Neerincx, M.A. (2003). Cognitive task analysis: Allocating tasks and designing support. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design*, pp. 283–305. Mahwah, NJ: Lawrence Erlbaum Associates.
- Nichols, S., Bristol, N. and Wilson, J.R. (2001). Workload assessment in railway control. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics*, Vol. 5: *Aerospace and Transportation Systems*, pp. 463–470. Aldershot, U.K.: Ashgate.
- Norman, D. and Bobrow, D. (1975). On data-limited and resource-limited processing. *Journal of Cognitive Psychology*, **7**, 44–60.
- North, P.A. and Riley, V.A. (1989). W/INDEX: A predictive model of operator workload. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton and L. van Breda (Eds.), *Applications of Human Performance Models to System Design*, pp. 81–89. New York: Plenum Press.
- ODCSPER. (1997). MANPRINT guidebook for systems design and assessment, July 1997. Personnel Technologies Directorate, ODCSPER. [http://www2.mitre.org/work/sepo/toolkits/risk/taxonomies/files/MANPRINT\\_Guidebook.DOC](http://www2.mitre.org/work/sepo/toolkits/risk/taxonomies/files/MANPRINT_Guidebook.DOC) (Accessed March 2014).
- O'Donnell, C.R.D. and Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman and J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, Vol. II, pp. 42/1–42/49. New York: Wiley and Sons.
- Paas, F., Tuovinen, J.E., Tabbers, H. and Van Gerven, P.W.M. (2003). Cognitive load measurement as means to advance cognitive load theory. *Educational Psychologist*, **38**(1), 63–71.
- Parks, D. and Boucek, G. (1989). Workload prediction, diagnosis, and continuing challenges. In G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton and L. van Breda (Eds.), *Applications of Human Performance Models to System Design*, pp. 47–64. New York: Plenum Press.
- Peck, E.M., Yuxsel, B.F., Ottley, A., Jacob, R.J. and Chang, R. (2013). Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proceedings of CHI 2013*. Paris, France: ACM.
- Peng, L. and Zhizhong, L. (2014). Comparison of task complexity measures for emergency operating procedures: Convergent validity and predictive validity. *Reliability Engineering and System Safety*, **121**, 289–293.
- Pickup, L., Wilson, J. and Lowe, E. (2010). The operational demand evaluation checklist (ODEC) of workload for railway signalling. *Applied Ergonomics*, **41**(3), 393–402.
- Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L. and Morrisroe, G. (2005a). The integrated workload scale (IWS): A new self-report tool to assess railway signaller workload. *Applied Ergonomics*, **36**(6), 681–693.
- Pickup, L., Wilson, J.R., Sharples, S., Norris, B., Clarke, T. and Young, M.S. (2005b). Fundamental examination of mental workload in the rail industry. *Theoretical Issues in Ergonomics Science*, **6**(6), 463–482.
- Pike, M., Maior, H.A., Porcheron, M., Sharples, S. and Wilson, M.L. (2014). Measuring the effect of think aloud protocols on workload using fNIRS. In *Proceedings of CHI 2014 ACM*, Toronto, Ontario, Canada, 26 April–1 May 2014.
- Recarte, M.Á., Pérez, E., Conchillo, Á. and Nunes, L.M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology*, **11**(2), 374–385.
- Reid, G.B. and Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*, pp. 185–218. Amsterdam, the Netherlands: North-Holland.

- Reimer, B. and Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: A field study and simulation validation. *Ergonomics*, **54**(10), 932–942.
- Richer, F. and Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, **22**, 204–207.
- Roscoe, A.H. (1987). In-flight assessment of workload using pilot ratings and heart rate. In A.H. Roscoe (Ed.), *The Practical Assessment of Pilot Workload*, AGARD-AG-282, pp. 78–82. AGARD, Neuilly-Sur-Seine, France.
- Rubenstein, J.S., Meyer, D.E. and Evans, J.E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, **27**, 763–797.
- Rubio, S., Díaz, E., Martín, J. and Puente, J.M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, **53**(1), 61–86.
- Saaty, T.L. (1980). *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill.
- Sarno, K.J. and Wickens, C.D. (1995). Role of multiple resources in predicting time-sharing efficiency: Evaluation of three workload models in a multiple-task setting. *International Journal of Aviation Psychology*, **5**, 107–130.
- Schneider, W. and Shiffrin, R.M. (1977). Controlled and automatic human information processing I: Detection, search, and attention. *Psychological Review*, **84**, 1–66.
- Schouten, J.F., Kalsbeek, J.W.H. and Leopold, F.F. (1962). On the evaluation of perceptual and mental load. *Ergonomics*, **5**, 251–260.
- Schulz, C.M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Wasmaier, M., Kocks, E.F. and Schneider, G. (2011). Eye tracking for assessment of workload: A pilot study in an anaesthesia simulator environment. *British Journal of Anaesthesia*, **106**(1), 44–50.
- See, J. and Vidulich, M.A. (1998). Computer modelling of operator mental workload and situational awareness in simulated air-to-ground combat: An assessment of predictive validity. *International Journal of Aviation Psychology*, **8**, 351–375.
- Senders, J. (1964). The human operator as a monitor and controller of multidegree systems. *IEEE Transactions on Human Factors in Electronics*, **HFE-5**, 2–6.
- Senders, J.W. and Posner, M. (1976). A queuing model of monitoring and supervisory control. In T.B. Sheridan and G. Johanssen (Eds.), *Monitoring Behavior and Supervisory Control*, pp. 245–259. New York: Plenum Press.
- Shannon, C.E. and Weaver, W. (1959). *The Mathematical Theory of Communication*. Urbana, IL: The University of Illinois Press.
- Sharples, S., Edwards, T., Balfe, N. and Wilson, J.R. (2012). Inferring cognitive state from observed behaviour. In N. Stanton (Ed.), *Advances in Human Aspects of Road and Rail Transportation*, pp. 633–642. Boca Raton: CRC Press.
- Sharples, S., Millen, L., Golightly, D. and Balfe, N. (2011). The impact of automation in rail signalling operations. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, **225**(2), 179–191.
- Shelton-Rayner, G.K., Mian, R., Chandler, S., Robertson, D. and Macdonald, D.W. (2012). Leukocyte responsiveness, a quantitative assay for subjective mental workload. *International Journal of Industrial Ergonomics*, **42**, 25–33.
- Shi, Y., Ruiz, N., Taib, R., Choi, E. and Chen, F. (2007). Galvanic skin response (GSR) as an index of cognitive load. In *Extended Abstracts CHI 2007*, pp. 2651–2656. San Jose, CA.
- Shingledecker, C.A. and Crabtree, M.S. (1982). Subsidiary radio communications tasks for workload assessment: II. Task sensitivity evaluation. Report No. AFAMRL-TR-82-57. Aerospace Medical Research Laboratory, Wright Patterson Airforce Base, OH.
- Shingledecker, C.A., Crabtree, M.S., Simons, J.C., Courtright, J.F. and O'Donnell, R.D. (1980). Subsidiary radio communications tasks for workload assessment in R&D simulations: I. Task development and workload scaling. Report No. AFAMRL-TR-80-126. Aerospace Medical Research Laboratory, Wright Patterson Airforce Base, OH.
- Smith, A. (1989). A review of the effects of noise on human performance. *Scandinavian Journal of Psychology*, **30**, 185–206.
- Tattersall, A.J. and Foord, P.S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, **39**, 740–748.
- Treisman, A.M. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97–136.
- Tsang, P.S. and Johnson, W.W. (1989). Cognitive demands in automation. *Aviation, Space, and Environmental Medicine*, **60**, 130–135.

- Tsang, P.S. and Velazquez, V.L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, **39**(3), 358–381.
- Vidulich, M.A. (1989). The use of judgement matrices in subjective workload assessment: The Subjective WORKload Dominance (SWORD) technique. In *Proceedings of the Human Factors Society 33rd Annual Meeting*, pp. 1406–1410. Santa Monica, CA: The Human Factors Society.
- Vidulich, M.A. and Bortolussi, M.R. (1988a). A dissociation of objective and subjective workload measures in assessing the impact of speech controls in advanced helicopters. In *Proceedings of the 32nd Annual Meeting of the Human Factors Society*, pp. 1471–1475. Santa Monica, CA: The Human Factors Society.
- Vidulich, M.A. and Bortolussi, M.R. (1988b). Speech recognition in advanced rotocraft: Using speech control to reduce manual control overload. In *Proceedings of the American Helicopter Society National Specialists' Meeting – Automation Applications in Rotocraft*. Atlanta, GA: Southeast Region of the American Helicopter Society.
- Vidulich, M.A. and Tsang, P.S. (1987). Absolute magnitude estimation and relative judgement approaches to subjective workload assessment. In *Proceedings of the Human Factors Society 31st Annual Meeting*, pp. 1057–1061. Santa Monica, CA: The Human Factors Society.
- Vidulich, M.A., Ward, G.F. and Schueren, J. (1991). Using the subjective workload dominance (SWORD) technique for projective workload assessment. *Human Factors*, **33**, 677–691.
- Warm, J.S., Dember, W.N. and Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman and M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications*, pp. 183–200. Mahwah, NJ: Lawrence Erlbaum Associates.
- Welford, A.T. (1968). *Fundamentals of Skill*. London, U.K.: Methuen.
- Wewerinke, P.H. (1974). Human operator workload for various control conditions. In *Proceedings of the 10th Annual Conference on Manual Control*, pp. 167–192. Wright-Patterson Air Force Base, OH.
- Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, **3**, 159–177.
- Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors*, **50**(3), 449–455.
- Wickens, C.D., Hollands, J.G., Banbury, S. and Parasuraman, R. (2013). *Engineering Psychology and Human Performance*, 4th edn. Boston, MA: Pearson.
- Wierwille, W.W. and Casali, J.G. (1983). A validated rating scale for global mental workload measurement applications. In *Proceedings of the Human Factors Society 27th Annual Meeting*, pp. 129–132. Santa Monica, CA: The Human Factors Society.
- Wilson, G.F. and Fullenkamp, P. (1991). A comparison of pilot and WSO workload during training missions using psychophysiological data. In *Proceedings of the Western European Association of Aviation*, Vol. II: *Stress and Error in Aviation*, pp. 27–34. Aldershot, U.K.: Avebury.
- Wilson, G.F., Purvis, B., Skelly, J., Fullenkamp, P. and Davis, I. (1987). Physiological data used to measure pilot workload in actual flight and simulator conditions. In *Proceedings of the Human Factors Society 31st Annual Meeting*, pp. 779–783. Santa Monica, CA: The Human Factors Society.
- Witchel, H.J., Westling, C., Needham, R., Healy, A. and Chockalingam, N. (2013). Mean head and shoulder heights when seated: Subconscious postural cycles during discrete, computerised stimuli. In *Proceedings of ECCE, 2013*, Toulouse, France.
- Wylie, G. and Allport, A. (2000). Task switching and the measurement of “switch costs”. *Psychological Research*, **63**, 212–233.
- Yeh, Y.Y. and Wickens, C.D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, **30**, 111–120.
- Yerkes, R.M. and Dodson, J.D. (1908). The relationship of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurological Psychology*, **18**, 459–482.
- Young, M.S. and Stanton, N.A. (1997). Automotive automation: Investigating the impact on drivers' mental workload. *International Journal of Cognitive Ergonomics*, **1**(4), 325–336.
- Young, M.S. and Stanton, N.A. (2002a). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors*, **44**(3), 365–375.
- Young, M.S. and Stanton, N.A. (2002b). It's all relative: Defining mental workload in the light of Annett's paper. *Ergonomics*, **45**(14), 1018–1020.
- Zheng, B., Cassera, M.A., Martinec, D.V., Spaun, G.O. and Swanström, L.L. (2010). Measuring mental workload during the performance of advanced laparoscopic tasks. *Surgical Endoscopy*, **24**(1), 45–50.
- Zwaga, H.J.G. (1973). Psychophysiological reactions to mental tasks. *Ergonomics*, **16**, 61–67.

---

# 19 Situation Awareness

*David Golightly*

## CONTENTS

|                                          |     |
|------------------------------------------|-----|
| Introduction.....                        | 549 |
| Theoretical Background.....              | 551 |
| Quantitative Measurement of SA.....      | 552 |
| Probes and Queries.....                  | 552 |
| Subjective (First-Person) Measures ..... | 554 |
| Observer (Third-Person) Measures .....   | 555 |
| Performance and Secondary Markers .....  | 556 |
| Elicitation and Modelling .....          | 556 |
| Expert and Participant Elicitation ..... | 556 |
| Event Analysis.....                      | 557 |
| Modelling and Representation .....       | 557 |
| Concluding Comments.....                 | 559 |
| References.....                          | 560 |

## INTRODUCTION

Situation awareness (SA) is the construct routinely applied to study of requisite awareness for dynamic tasks. With roots in aviation (see van Winsen et al., 2014, for a historical analysis of the origins of SA), SA has been widely studied in domains as diverse as nuclear control (e.g. Hogg et al., 1995; Hallbert, 1997), train driving (e.g. Tschirner et al., 2013) and rail maintenance (e.g. Golightly et al., 2013), off-shore engineering (e.g. Sneddon et al., 2006), power supply and generation (e.g. Salmon et al., 2008), driving (e.g. Ma and Kaber, 2005), surgery and anaesthesia (e.g. Fioratou et al., 2010) and space operations (e.g. Chiappe et al., 2014). Understanding what constitutes SA in these domains is believed to play a key role in shaping requirements and evaluation criteria for new technology, user interfaces and automation, as well as for new procedures, such as protocols for safety-critical communications, and organisational structures. It may also be used to derive requirements for training and assessment, particularly in the area of non-technical skills, for selection criteria and as a framework for, or component of, incident analysis. SA also has a reach beyond traditional human factors fields into topics such as reading, chess and sport (Durso and Dattel, 2004; Bourbousson et al., 2011).

Though intuition does not by necessity make a sound theoretical platform (Flach, 1995; Dekker and Hollnagel, 2004), part of the appeal and acceptance of the SA concept is no doubt due to an intuitive sense that understanding what is going on around oneself is important to effective decision making and action, and operators in various domains naturally talk about the importance of having *the picture* (air traffic control), *the plot* (rail control) or *the bubble* (naval operations). Another reason for the durability of SA as a construct is that it has resulted in a suite of methods. Many of these take the form of quantitative measures that aim to capture some reflection of the currency or accuracy of an individual or team's SA, or some first- or third-party subjective evaluation of the state of SA. While less commonly classified as part of the ergonomics/human factors (E/HF)



**TABLE 19.1**  
**Categorisation of SA Methods**

| Quantitative              |                                                              | Elicitation and Modelling          |                                                             |
|---------------------------|--------------------------------------------------------------|------------------------------------|-------------------------------------------------------------|
| Category                  | Examples                                                     | Category                           | Examples                                                    |
| Probe                     | SAGAT (Endsley, 1995b)                                       | Expert and participant elicitation | Retrospective video analysis (e.g. Bourbosson et al., 2012) |
|                           | SPAM (Durso and Dattel, 2004)                                |                                    | Protocol analysis (e.g. Salmon et al., 2013)                |
|                           | SACRI (Hogg et al., 1995)                                    |                                    | Critical decision method (Crichton and Flin, 2004).         |
| Subjective                | SART (Taylor, 1990)                                          | Event analysis                     | Single event (e.g. Gorman et al., 2006)                     |
|                           | SARS (Bell and Waag, 1997)                                   |                                    | Multiple event (e.g. Rogers et al., 2000)                   |
| Objective (third person)  | SARS (Bell and Waag, 1997, as reported in Saus et al., 2006) | Informal modelling                 | Narrative (e.g. Artman and Garbis, 1998)                    |
|                           | SABARS (Matthews et al., 2000)                               |                                    | Knowledge elements (e.g. Bourbosson et al., 2012)           |
|                           | ANTS (Fletcher et al., 2003)                                 |                                    | Functional analyses (e.g. Golightly et al., 2013)           |
| Performance and secondary | Readback errors (e.g. Durso et al., 2006)                    | Formal modelling                   | Goal-directed task analysis (Endsley and Rogers, 1994)      |
|                           | Eye tracking (e.g. Underwood et al., 2002)                   |                                    | WESTT (Houghton et al., 2008)                               |
|                           | EEG (e.g. Saus et al., 2006)                                 |                                    | EAST (Walker et al., 2006)                                  |

researcher's toolkit for the study of SA, there is also a role for elicitation and modelling approaches. These may be used as task analyses to inform the design of quantitative tools but are also regarded as valuable in their own right.

The aim of this chapter is to map out the landscape of methods explicitly associated with SA, as summarised in Table 19.1. The emphasis on *explicit* association is important as the potential centrality of SA (or, its critics might argue, its ill-defined nature) within cognition and performance mean that many other methods and measures may be studying aspects related to SA, such as workload (Hart and Staveland, 1988) or team mental models (Langan-Fox et al., 2000). In order to bound this chapter, the criteria for inclusion of a method are that it either specifically sets out to address SA or that one of its outputs are data that have been explicitly linked to SA, for example, when the use of critical decision method (a more general method) has highlighted the importance of SA in incident control (Crichton and Flin, 2004). Given the number of methods available (Stanton et al. [2005] identify over 30 SA quantitative methods alone), this review is not intended to be exhaustive, but instead demonstrates the major classes of methodology, and some of the underpinning rationale behind these classes, in order to support an informed decision about how one might proceed with an evidence-based investigation of SA.

The chapter is structured as follows. Due to the diversity of approaches and the theoretical positions they reflect, the chapter first covers some of the underpinning theory, common debates and resulting implications for methods. More extensive theoretical analysis can be found in edited volumes by Endsley and Garland (2000) and Banbury and Tremblay (2004) and the landmark special issue of Human Factors (Gilson, 1995). This theory should also serve as the basis for the interpretation of the results of any enquiry into situation awareness. This is followed by presentation of quantitative measures, including some of the studies to compare the outputs of measures, which have historically shaped the study of SA. Methods are then described relating to elicitation and/or modelling of domains and domain characteristics (see Chapter 7 for further discussion of elicitation



tools) and the chapter concludes with thoughts about future directions for the application of methods relevant to SA.

## THEORETICAL BACKGROUND

While there are numerous definitions of SA, these definitions show consistencies. Many definitions refer to both comprehension of current state and prediction into the future. For example, the most common formulation of SA (Endsley, 1995a) comprises three levels – perception of the situation, comprehension of the situation and prediction of future events. Therefore, SA is as much about anticipating future events as it is about interpreting the current state of the world. Endsley's approach to SA, and others like it, comprise a Cartesian (Dekker, 2013) or knowledge-based (Gorman et al., 2006) approach to SA, with an emphasis on SA being *knowledge of the world*, as distinct from the world itself. As we shall see, this orientation underpins many measurement approaches to SA such as Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1995b) and the Situation Present Assessment Method (SPAM) (Durso and Dattel, 2004) which query an operator's understanding of unfolding events at a given time.

There are complexities with this description of SA. For example, rather than experts having knowledge of both current and future states, there is evidence that their knowledge may be predominantly future orientated and that knowledge of current states may be less relevant to performance (Durso et al., 1998). Also, rather than recalling detail, experts recall the *gist* of information, based on groups of elements that are particularly salient at any one time, such as the relative altitudes between two aircraft at risk of collision (Gronlund et al., 1998; Hauss and Eyferth, 2003). This kind of prioritisation and perceptual grouping is in line with historical work on expertise (e.g. Chase and Simon, 1973; Chi et al., 1981) and therefore must be considered when taking into account the elements, clusters of elements and element-related characteristics that might be relevant to an individual, particularly an expert operator, at any given time.

As well as the *components* of awareness being critical in dynamic environments such as the flight deck or process control, the *temporal* dimension is also important (Sarter and Woods, 1991). System state and priorities emerge over time, and failures and emergencies potentially result from a series of minor failures compounding and escalating in severity rather than one key event. The capacity of working memory is such that it may not, alone, be able to track the volume of variables in an unfolding, dynamic picture, even with priority-based chunking (Hauss and Eyferth, 2003). There is also evidence that knowledge of future states for highly complex, non-transparent systems may be very difficult to report, even though operators are able to demonstrate effective performance suggestive of implicit knowledge of those systems (Berry and Broadbent, 1988; Gorman et al., 2006).

Theoretically, the contention is that awareness is an ongoing and iterative process of perception, interpretation and action (Neisser, 1976; Smith and Hancock, 1995) that calls into question the requirement for, or even the cognitive viability of, a full symbolic representation of the world for the operator as awareness. This position contends that SA exists as *externally directed consciousness* (Smith and Hancock, 1995), where the environment (the dynamic task elements, the physical environment, the displays and artefacts that represent the task) is more than mere context that feeds into information processing, but the ground in which a cycle of awareness, interpretation and resulting performance, is formed (Artman and Garbis, 1998). Appropriate (or, in adverse situations, inappropriate) schemata (broadly defined as knowledge templates for interpretation and action [Alba and Hasher, 1983]) are recalled from long-term memory to support the cycle (Stanton et al., 2009) and shape the sampling of the world as well as supporting the interpretation of data from the world. It is important to stress this recall of schemata is neither conscious nor voluntary, but rapid and perceptually linked to events.

The term distributed SA has been adopted to describe the interpretation of SA as something that is not purely cognitive or a product of internal representation, but constructed across both the individual and the environment. This approach has been most prevalent in (but is not exclusive to) team

contexts, where the awareness necessary for appropriate performance emerges as an active process of actors (human and automated) working together (Chaippe et al., 2012). The most extreme view is that distributed SA may not be reducible to the awareness of individuals, but can only be understood as an emergent property of the system as a whole system including people, artefacts and environment (Salmon et al., 2008; Stanton et al., 2009).

This debate has implications for methods. First, if SA is distributed to a greater or lesser extent, then the environment or context must be in place, and its current status actively considered by the investigator, during any testing or examination of SA. This has led both to probe techniques such as SPAM (Durso and Dattel, 2004) where the display is active and visible during queries, as well as interest in techniques that model the world and interactions within the world, rather than merely probe knowledge. Second, in team settings, if the shared SA approach is essentially an extension of the knowledge-based approach, team SA can therefore be effectively studied by using related methods, such as SAGAT, applied at a team level. If, however, distributed SA is to a greater or lesser extent an emergent property, then the method needs to take a broader stance and examine networks of information flow across the team in response to various roadblocks or changes to the expected norm (Gorman et al., 2006), and how this might vary due to task or role (Salmon et al., 2008; Houghton et al., 2008).

There are a few things that situation awareness is *not*. It is not performance – it is possible to have complete SA, but then take suboptimal decisions, or commit errors in planning and action. Similarly, it is possible to have incomplete SA, but through strategy, skill and maybe plain good fortune, demonstrate adequate performance (Durso et al., 2006). The implication for methodology therefore is that explicit measures of performance are not always reliable indicators of the state of SA. That said, some of the evidence for the reliability, validity and sensitivity of SA measures such as SAGAT and SPAM has come from contrasting measures of SA with performance. For example, Paige Bacon and Strybel (2013) found in a simulated Air Traffic Control (ATC) task that loss of separation events and lower values of lateral separation were predicted by scores on SA measures. Also, SA is not workload, though there is a relationship between the two and while the effort to maintain SA may be substantial (Adams et al., 1995), methods to measure SA should be distinct and measure a different construct from workload tools. Finally, while some have questioned whether SA is merely a catch-all term rather than a unique construct in its own right (Flach, 1995; Bell and Lyon, 2000), there is evidence that measures of SA and, by extension the construct of SA, explains dynamic performance in a way that measures of underpinning cognitive mechanisms cannot explain on their own (Durso et al., 2006). There is a particularly close relationship with mental models, in that an effective mental model may be required to underpin accurate interpretation of the current and future state of the system (Sarter and Woods, 1991). The two, however, are distinct in that the mental model can be used for planning, potentially through mental simulation, whereas SA is based on the current (and potential future) reality of an occurring situation (Hrebec and Stiber, 2001).

It is worth stressing again that as well as highlighting the origins of the various methods covered in the rest of this chapter, this brief theoretical analysis highlights the assumptions that an investigator is committing to when adopting a particular methodological approach. This influences not only the choice of method, but the stance the investigator should take when interpreting the data that a given method delivers. Having presented this theoretical background, and the ongoing debate as to the nature of SA, we now turn to methods themselves.

## QUANTITATIVE MEASUREMENT OF SA

### PROBES AND QUERIES

The first group of methods presented originates from the position that SA is testable and measurable, and therefore the methods take the form of quantitative measures. The predominant approach to measuring SA is through the probe technique. Operator(s) are asked a series of questions, during

some kind of real-time or simulated operation of a system, in order to ascertain the current state of knowledge or understanding of the system. Among these methods, SAGAT is the most widely used to date (Endsley, 1995b).

SAGAT involves deriving a set of queries based around the investigator's assessment of the SA requirements of any given domain. In accordance with the Endsley three-level model, queries are designed to probe perception, current comprehension and prediction. While *best practice* is to derive these queries through a goal-directed task analysis (discussed later), the queries may also be derived directly from more general task analyses or from subject matter experts' (SMEs) impressions of the most relevant queries for a given domain. The study participant is then required to work within a given setting. At various task in interrupted at some point, the participant either turns away from the workstation or displays and relevant artefacts are blanked or removed, and one or more queries are presented. SAGAT can use free recall, but has also been applied with cued recall allowing for a higher number of queries to be answered in a short space of time. After a set of queries is complete, the participant may resume where they left current task or move to a new scenario. This approach generates a general SA score, as well as specific scores applicable to the three levels. See Kaber and Endsley, 2004, for an application of SAGAT. SAGAT has also been applied in team settings. For example, Bolstad and Endsley (2003) report SAGAT's capacity to detect differences between functional units within army operations and differences between unit commanders and front-line personnel.

SAGAT is the most common exemplar of a general form of query method. For example, specific forms of probe representation might be used, such as presenting operators with a map or schematic where they can position elements such as trains for rail signallers (Golightly et al., 2009) or planes for ATCOs (Gronlund et al., 1998). Also, probe questions about domain elements need not necessarily be structured or analysed around the Endsley three-level model. For example, Gronlund et al. (1998) categorised probes with reference to an aircraft's risk of loss of separation at the time of the freeze, successfully identifying varying levels of knowledge with respect to aircraft in an ATC simulation. Another iteration of SAGAT is the Situation Awareness Control Room Inventory (SACRI) tool (Hogg et al., 1995). Not only does SACRI specify a question set for use in a nuclear control setting, it adapts the scoring of a standard SAGAT by using signal detection theory, for example scoring on false positives. SACRI has been shown to be sensitive to competency, simulated disturbances, and HMI.

Despite its widespread use, there are a number of potential drawbacks of SAGAT. One concern is that the interruption and removal of the scenario from the operator during freezes make SAGAT primarily a test of working memory. SAGAT scores may therefore be a reflection of general individual differences in working memory (Just and Carpenter, 1992), or more specific strategic differences in use of memory exhibited by operators, for example, in how rail signallers vary in their strategic style and memory of train position during task freezes (Golightly et al., 2009).

More broadly, there is the theoretical concern that SA is not purely a product of cognition, but is developed between the individual and the external world including the display, notes, memory aids and, in team settings, other individuals. SPAM (Durso and Dattel, 2004) attempts to resolve this issue of removing the domain by keeping the dynamic situation running and present in front of the participant. It uses query latency, rather than query accuracy, as the main dependent variable. Therefore, the display plays an important role in providing real-time information on SA. This use of the probe in real time and with the display present also potentially allows the method to be used in the field rather than purely in the simulator. Because it uses latency as a metric, SPAM supports measures of when SA is successful, not just when it fails. Typically the request is presented as a phone call from the experimenter, therefore the length of time taken to answer the call or whether the participant is able to answer the call at all is purported to be indicator of workload (Durso and Dattel, 2004), though there is evidence that questions the reliability of this measure when response time is compared with dedicated measures of workload (Strybel et al., 2011). Vu et al. (2010) give an example application of the SPAM approach.

Similar to queries at three different levels used in SAGAT, SPAM queries are designed to distinguish between understanding of current system state, and understanding of future system state. This distinction has been used successfully in the past to demonstrate differences between experts, who typically score better on future-orientated queries in comparison to novices (Durso and Dattel, 2004). Durso et al. (2006) compared SPAM and a version of SAGAT, finding that the two measures showed a high degree of correlation. However, when these measures were added to other cognitive measures as a model to explain performance on a simulated ATC task, the predictive power of SPAM was considerably greater in comparison to that of the SAGAT-like probe. The explanation offered was that the variance attributable to the SAGAT measure had already been accounted for by measures of working memory, raising again the issue of whether SAGAT is principally a measure of working memory, rather than a specific measure of an independent SA construct.

One of the most critical aspects of any probe method is the validity of the probes used. Several applications base probes on a combination of experience of the investigator, and of querying elements that are of particular relevance at any time (e.g. identifying two planes that may be at risk of loss of separation). Indeed, once such constraints have been applied, the number of questions that might be relevant and meaningful at any point in time might be extremely limited (Hauss and Eyferth, 2003; Golightly et al., 2012). It may be useful to retrospectively assess content validity with participants, as was done with the query set used in SACRI (Hogg et al., 1995). Hauss and Eyferth (2003) developed SALSA ('measuring situation awareness of area controllers within the context of automation' in its original German), which takes a SAGAT-like approach to ATC SA, but then weighs elements based on retrospective expert analyses of the scenario. The results of SALSA found that recall performance was higher when taking into account weighting of priority than recall performance using standard SAGAT. Similarly, Kaber et al. (2006) used a weighted SAGAT based on data collected in real time during the simulation to discriminate between different levels of SA as defined in the Endsley (1995a) model for adaptive automation applied at different stages of information processing, and Jeannot et al. (2003) developed SASHA-L (SA for SHAPE On-Line; where SHAPE is *Solutions for Human-Automation Partnerships in European ATM*), a derivation of SPAM where SMEs ask questions in real time based on current system state, and providing a ranking of importance at that time. Salmon et al. (2009) conclude that probe measures may be most effective when 'SA-related elements and associated states and properties can be accurately identified prior to task performance' (p. 498) – in other words, when contents, and crucially the timing (Golightly et al., 2012) of probes, can be determined by using a preplanned, simulated scenario.

There is also a concern with probe techniques that the presence of queries may interfere with the performance of tasks and of SA itself either by positively highlighting salient elements of a task or negatively by interrupting the operator. However, evaluations of both SAGAT and SPAM-like measures have not found evidence of interference (Endsley, 1995b; Durso et al., 2006; Paige Bacon and Strybel, 2013).

## SUBJECTIVE (FIRST-PERSON) MEASURES

Subjective first-person approaches use some quantification of self-report on the part of the participant or operator to measure SA. This family of measures stems from the view that SA is explicitly available for self report, and typically use multiple scales to capture SA ratings (Jones, 2000).

The most common of the subjective measures is the Situation Awareness Rating Technique (SART) (Taylor, 1990). SART uses a 10-item scale to capture subjective ratings over three major dimensions – task demand, supply of attentional resources and understanding. Calculating scores for the higher level dimensions is intended to be used in an algorithm ( $SA = \text{Understanding} - [\text{Demand} - \text{Supply}]$ ) to give an overall score of SA. A simplified version of SART – 3D SART (Jones, 2000) – condenses the 10 points down to just three scales corresponding to the three major dimensions. Other subjective measures include SA-SWORD (SA Subjective WORKload Dominance) technique (Vidulich and Hughes, 1991), which is used in a within-subjects design to compare SA

across different system designs. One interesting alternative to generic subjective measures is SARS (SA rating scales) (Bell and Waag, 1997), which uses scale items that are not generic but are based on specific task characteristics (military aviation), though it takes the view of SA as an ability or personal characteristic, rather than as a changeable state determined by task factors.

Empirical support for subjective measures has been equivocal. On the one hand, there is evidence that subjective measures can discriminate between different tasks characteristics such as display (Taylor and Selcon, 1990). However, correlation between SART and other measures of SA such as SAGAT (Jones and Endsley, 2000; Salmon et al., 2009) and SPAM-like measures (Golightly et al., 2012) have typically been more elusive. Examination of subscales has identified that changes in independent variables, such as display design, may lead to the experiences of changes in demand rather than SA per se (Jones, 2000), and despite promise for SARS as a subjective technique (Jones, 2000), a modification for naval operations failed to correlate with performance (Saus et al., 2006) (further discussion of the demand/effect debate is presented in Chapter 18). One concern is that while there is an argument that subjective sense of SA is important, particularly when evaluating new technology or procedure to ensure operators believe it is providing them with the level of information and support they require, there is nothing to say this is correlated with onward performance (Fracker and Vidulich, 1991). That said, a correlation between performance and a subjective sense of awareness may be equally problematic as participants will use perceptions and measures of performance as a valuable source of feedback as to how much they know about the current task. In addition, the participant will, by definition, be mostly unaware of what they do not have awareness of, and therefore be unable to give accurate assessment of their levels of awareness. Overall, the value of subjective measures of SA continues to be questioned and the methods have so far only found limited support.

### **OBSERVER (THIRD-PERSON) MEASURES**

Third-person measures involve an observer scoring SA characteristics of an operator or participant. Observer measures can be used during live performance, or through retrospective analysis, for example reviewing video. While other measures may involve qualitative review of a given domain, observer measures specifically try to score or count general types of behaviour or strategy, or specific behavioural markers.

SARS (Bell and Lyon, 2000) has been employed for use by third-person expert reviewers and peers, demonstrating a high degree of correlation between those SA-relevant characteristics measured in SARS, and measures such as errors and communications. Also, a derivation of SARS that failed to correlate with performance measures in first-person form was successful as a measure in a modified form for third-person observers (Saus et al., 2006). While the measures in SARS are based on characteristics, or classes, of activities (e.g. *threat prioritisation*, *defensive reaction*), other tools have identified highly specific behaviours that are believed to be indicative of SA. One such example is Situation Awareness Behaviourally Anchored Rating Scale (SABARS) (Matthews et al., 2000) where 28 behavioural markers were identified by experts (e.g. *sets appropriate levels of alert*, *solicits information from squad leaders*), which can then be scored by observers. Typically, these observer measures are used in training and selection settings, and may be used in conjunction with other measures of performance. They are of particular relevance when it comes to training and assessing non-technical skills, of which SA is considered to be critical in many domains. For example, the ANTS tool (Anaesthesia Non-Technical Skills) (Fletcher et al., 2003) captured eight of the non-technical skills relevant to anaesthesiology. This included situation awareness which, like the other skills in ANTS, was broken down into a number of high-level characteristics with a series of behavioural markers for good and poor practice. These markers could then be identified and scored during observation sessions.

One noted issue with observer measures is that it may be difficult to apply observer measures in team settings where members are not co-located due to the need to find a number of suitably



experienced observers (Salmon et al., 2009). Also, while such measures are useful for third-person observation, they do not necessarily translate into first-person measures for self-assessment of SA. While SARS has been used in both subjective and objective settings at least for military aviation, the same transfer has not been demonstrated for SABARS (Matthews et al., 2005).

## PERFORMANCE AND SECONDARY MARKERS

While being aware that SA is not directly related to overall task performance – good awareness of what is occurring does not always lead to effective action (Endsley and Garland, 2000) – specific, embedded performance measures could be used to assess the point at which emerging task factors become salient to operators in simulated tasks (Sarter and Woods, 1991). An example of this in practice is identification of a secondary event, such as identifying a readback error (Durso et al., 2006) in ATC simulation. Bell and Lyon (2000) use both communication type and count, in addition to error count, to demonstrate correlations between observable behaviours and observed SARS scores.

Psychophysiological measures may also have a place in the study of SA. Eye tracking (Smolensky, 1993) has been applied, particularly for SA in driving (e.g. Underwood et al., 2002) while physiological measures, such as EEG, blink rate and duration and heart rate variability (Saus et al., 2012), have also achieved some level of success as an index of SA. The drawback of such approaches is that they are difficult to achieve outside of the laboratory and, in comparison to some of the other measures discussed here, complex to analyse. However, they have the benefit of providing continuous data, in comparison to probe measures, often at a level of sensitivity above that of rating measures (Wilson, 2000).

## ELICITATION AND MODELLING

The target of SA elicitation approaches is to identify the characteristics of the environment, task or activity that shape performance where awareness is utilised. Models typically express characteristics such as task models (Endsley and Rogers, 1994), performance constraints (Smith and Hancock, 1995) or networks of communication (e.g. Salmon et al., 2008). Elicitation and modelling techniques serve a dual purpose in SA methodology. Elicitation is often the first step to capturing SA requirements for queries to be used in probe-based approach, discussed in the previous section. However, for some, elicitation and modelling is a goal in its own right that seeks to move away from purely cognitive, Cartesian perspectives, particularly for team and collaborative settings (Chiappe et al., 2014), aligning these methods with the theoretical position of awareness as an active process between individuals and/or teams, and the environment in which they are working.

Several of the methods that fall under elicitation and modelling are already well known (e.g. Critical Decision Method [CDM], hierarchical task analysis; see Chapters 3 and 6 of this book) but have been explicitly applied to the study of situation awareness. The aim here is not to fully describe them, but instead to demonstrate their relevance to the study of SA.

## EXPERT AND PARTICIPANT ELICITATION

Many of the conventional techniques for real-time elicitation, such as observation and ethnography, are used in contexts to capture the characteristics of SA and attention in dynamic settings. The risk is that a degree of information about what constitutes SA in such contexts remains tacit. Therefore, it is often advisable to complement observational data with additional elicitation from participants. Online elicitation may be most effective in tasks with a single operator, and where there is little risk of performance impact due to concurrent verbalisation. For many tasks, this is not possible and so offline elicitation is preferable. This may be carried out by asking participants to verbalise their processes and knowledge while watching a video of their initial task performance



(e.g. Bourbousson et al., 2011). Real-time protocols can also be analysed afterwards (see Chapter 7). For example, Salmon et al. (2013) used a number of metrics described from Event Analysis of Systemic Teamwork (EAST) (Walker et al., 2006; see later section) to compare the breadth of comments from both novice and expert drivers on naturalistic drives over level crossings, determining strategic differences in expertise that can be associated with SA. CDM (Klein et al., 1989) may also be used to elicit the cues relevant to understanding or interpreting incidents within a given domain, for example, by running experts through a specific command and control scenario (Salmon et al., 2008). Similar approaches have also been used in the understanding of non-technical skills, including those relevant to situation awareness (Crichton and Flin, 2004).

## EVENT ANALYSIS

Event analysis involves investigation and interpretation of events, usually some kind of incident, accident or near miss, retrospectively by a third party to ascertain either the general characteristics of the event (of which SA may turn out to be a contributory factor) or specifically with a view to assessing SA. We distinguish these from some of the offline elicitation approaches in that the events are often reconstructed from archival sources such as incident reports rather than directly using the actors involved. These can be further subdivided into single event and multiple event.

Single event is where one event is inspected in great detail to elicit events that lead to SA problems, such as Gorman et al.'s (2006) detailed examination of a friendly fire incident or Dekker and Lutzhoft's (2004) examination of a cruise ship grounding. This kind of single-event analysis can be useful as an inductive approach to understand either the nature of SA within a very specific situation (and given arguments for the task-specific nature of SA, the details of any particular event or its precursors play a critical role) or, as in the case of Dekker and Lutzhoft (2004), to propose alternative interpretations to a conventional SA approach.

Multiple event analyses are performed with the use of a theoretical framework to analyse events in terms of their SA components and characteristics. For example, Jones and Endsley (1996) reviewed a body of ATC-reported incidents and near misses to identify the proportion that could be attributed to issues with SA. Furthermore, they then classified these incidents in terms of whether the incidents were primarily due to failures at levels 1, 2 and 3, finding an overwhelming proportion of incidents attributable to issues associated with initial perception (level 1). This approach has also been applied to offshore oil and gas operations (Sneddon et al., 2006). As well as using SA levels as a categorisation of incidents, another approach is to classify the task and environmental events that have influenced SA and potentially led to performance issues and incidents. For example, Rogers et al. (2000) classify ATC events not in terms of SA levels, but by the kind of flight paths in relation to sectors that resulted in degraded SA.

In sum, the event analysis approach can lead to both microscopic and macroscopic interpretation of SA in relation to task characteristics, depending on whether a single- or multiple event approach is used. However, it is in the arena of event analysis that the caution expressed by Dekker (2013) must be taken most seriously. SA in event analysis can be used as an analytical construct, but this should not slide into SA, and specifically '*flaws or errors*' in an operator's SA, being attributed as a root cause in accident investigation.

## MODELLING AND REPRESENTATION

The kind of elicitation approaches described earlier will often result in some attempt to model or structure the characteristics of SA. That said, it is not always the case that there is a desire to formally model the outputs of SA-orientated elicitation. Instead, investigators may choose to present their data as narratives, especially when using single-case retrospective analysis. This approach often reflects a desire on the part of the author's to demonstrate the variety of richness of the data and may reflect a conscious desire to avoid narrowing SA down to one or more defined elements

are processes. For example, Dekker and Lutzhoft (2004) argue for the limitations of a purely SA-orientated analysis through the use of a single narrative account the factors and events leading up to a cruise ship grounding. Similarly, Garbis and Artman (2004) use a qualitative, descriptive account to describe the critical role of processes, teams and artefacts in SA as part of distributed control operations.

With that noted, it is often useful to structure, if not strictly model, qualitative data. For example, Bourbousson et al. (2011) abstract tables of shared knowledge elements from their analysis of a team sport. Golightly et al. (2013) use tables of functional analyses to highlight the role of situation awareness in protection and isolation for rail engineering. Blandford and Wong (2004) use a combination of floor plan schematics, models of communication flows and classification of knowledge against the Endsley three-level model, in addition to quotes and more narrative accounts, to describe the nature SA in an emergency dispatch control centre. Also, the purpose of the SA analysis may dictate the form of presentation. Tschirner et al. (2013) use a combination of real-time and offline approaches to study SA for the purposes of designing driver advisory systems for the railway. As a result, their output is presented as a set of textual requirements for a future system.

In many instances, there are specific attempts to model the domain in a formal manner. In some cases, this approach involves referencing SA-relevant aspects of a domain within a larger framework. For example, a Cognitive Work Analysis (CWA) (Vicente, 1999) may make reference to *building and maintaining SA* as an abstract function within a work domain analysis (Bisantz and Masaeva, 2009), which is then linked to underlying physical functions and forms.

Modelling approaches may also be used to express the tasks carried out within the domain of interest. While more general hierarchical task analysis has been used to this end (Salmon et al., 2009), one specific form is the goal-directed task analysis. The goal-directed task analysis (Endsley and Rogers, 1994) has been linked specifically to the derivation of probe cues for SAGAT. GDTA supports the elicitation of the goals of any tasks, and therefore the information necessary to achieve those goals. From here it is possible to identify the aspects of any given domain that need to be monitored for effective performance. Typically, this type of qualitative elicitation is the first step towards developing content for a measure such as SPAM or SAGAT queries. For example, Endsley (1993) used a combination of interviews and goal-directed task analysis to derive a list of dynamic elements relevant to the air-to-air combat domain. From there, a survey tool was developed with participants ranking the importance of each element, leading to a requirements analysis and information on priority among elements for that domain. In practice, any of the observational and elicitation tools described here could be used to develop probes. For example, Golightly et al. (2010) used prolonged observation with concurrent commentary with signallers to develop tools for assessing signaller/train dispatch situation awareness.

There are also a number of modelling approaches that provide specific representations of the domain, relevant to SA, even if not all were designed with that use in mind. Reflecting a theoretical perspective that team SA is derived from the interactions of team members, the Coordinated Awareness of Situations in Teams (CAST) (Gorman et al., 2006) approach uses an analysis of events, termed roadblocks, which threaten or actually damage team performance. The approach then examines both individual and team SA requirements with an emphasis on communication and coordination within the team, resulting from these roadblocks. This approach has been used to highlight the requirements, and potential deficiencies, of differentiated perspectives within the team, and the need for coordinated action and information exchange for successful team performance.

Also, modelling of the elements required for SA awareness may effectively form an ontology of domain elements. These may vary from the relatively informal in that they do not impose any particular relationships between the elements expressed, such as the structuring of domain elements relevant to air traffic control (Ruitenbergh, 1997; Jeannot et al., 2003), through to those that more formally express the relationships between elements, such as *is a* or *has a* relationships. A good

example of this is the knowledge network derived as part of the Workload, Error, Situational awareness, Time and Teamwork (WESTT) methodology (Houghton et al., 2008). Based on data derived from sources such as CDM, WESTT presents three representational forms based around task structure, social network and knowledge network. This knowledge network represents key points of information that are exchanged during collaborative activities. By combining the three representations, quantitative measures of task characteristics, such as SA requirements for different task phases, can be derived. Sociometric analysis can also give an indication of the centrality of any particular piece of information. In keeping with the distributed SA perspective, this supports an analysis of where awareness lies at any one time across the system as a whole. Similarly the EAST framework (Stanton et al., 2006) integrates a hierarchical task analysis, a coordination demand analysis, a communications usage diagram, a social network analysis and the critical decision method to provide two summary representations in the form of an enhanced operation sequence diagram and a propositional network (see Salmon et al., 2008 for an application of EAST).

This final point illustrates a general theme of elicitation and modelling approaches. In practice, methods need to be used in combination to triangulate data (Dekker, 2000 and Chapter 1 of this book). For example, Bourbousson et al. (2011) use *course-of-action* analysis involving video analysis of a team activity (a basketball team, in their case) followed by interviews with team members including retrospective commentary on the videos. The subsequent modelling approach that one may elect to take is as likely to reflect the practical goals of a given SA investigation as it does any particular theoretical perspective.

## CONCLUDING COMMENTS

As an overview of SA methods, while the probe measures of individual awareness still have a prominent role in the literature, they are best served for laboratory/simulation applications. Also, probes (and investigators' knowledge of how to apply the probes) must be highly domain specific, not just in their probe content, but how they are timed and how they reflect the current priorities. Any investigator wishing to use such an approach must therefore not underestimate the effort required to fully understand the domain first. The evidence is that probe measures are best suited to individual activities, that SAGAT and SPAM are most prevalent in the literature, but also that there is value in some of the less well-known techniques (e.g. SACRI) to tease out factors such as probe relevance at any one time. Subjective measures are less prominent overall in the literature and seem to be of most value when structured in terms of highly specific task elements. However, task elicitation approaches are of increasing interest, particularly as SA attempts to tackle complex, collaborative tasks and domains. Their very ethos is to capture domain-specific characteristics and represent an increasing acknowledgement of SA as an emergent property of people, task and artefacts, rather than a purely cognitive phenomenon. It seems likely that the prominence of these methods for the study of SA will only grow.

Any method comes with underpinning assumptions on theory and by adopting a method we adopt a theoretical position. This is nowhere more apparent than in the study of SA, where the breadth of methods that have been described in this chapter are just one manifestation of a deeper debate as to the nature and relevance of SA. While there are strong arguments to accept SA as a valid area (Parasuraman et al., 2008), it is wise to heed the caution expressed in the critiques, particularly the risk of applying the construct SA as a causal factor for events such as catastrophic failures and therefore, by extension, as some kind of error on the part of an operator, when such system failures are not reducible to an individual's state of knowledge but part of an ongoing set of design and situational features (Dekker, 2013). Therefore, it is an ethical responsibility of the human factors investigator to use these methods in a manner that seeks to understand onward causal factors and understand theoretical mechanisms for more effective future work, process and technology design.

## REFERENCES

- Adams, M. J., Tenney, Y. J. and Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 85–104.
- Alba, J. W. and Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, 93(2), 203.
- Artman, H. and Garbis, C. (1998). Team communication and coordination as distributed cognition. In *Proceedings of Ninth Conference of Cognitive Ergonomics: Cognition and Cooperation* (pp. 151–156).
- Banbury, S. and Tremblay, S. (Eds.). (2004). *A Cognitive Approach to Situation Awareness: Theory and Application*. Hove, UK.
- Bell, H. H. and Lyon, D. R. (2000). Using observer ratings to assess situation awareness. In M. R. Endsley (Ed.), *Situation Awareness Analysis and Measurement*. Mahwah, NJ: Laurence Erlbaum Associates.
- Bell, H. H. and Waag, W. L. (1997). *Using Observer Ratings to Assess Situational Awareness in Tactical Air Environments* (No. AL/HR-TP-1996-0050). Mesa, AZ: Human Resources Directorate, Aircrew Training Research Division, U.S. Air Force, Armstrong Laboratory.
- Berry, D. C. and Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, 79(2), 251–272.
- Bisantz, A. M. and Mazaeva, N. (2009). Work domain analysis using the abstraction hierarchy: Two contrasting cases. In: A. M. Bisantz and C.M. Burns (Eds.), (2008). *Applications of Cognitive Work Analysis*. (pp. 15–47). Boca Raton: CRC Press.
- Blandford, A. and William Wong, B. L. (2004). Situation awareness in emergency medical dispatch. *International Journal of Human-Computer Studies*, 61(4), 421–452.
- Bolstad, C. A. and Endsley, M. R. (October 2003). Measuring shared and team situation awareness in the army's future objective force. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 3, pp. 369–373). Santa Monica, CA: Human Factors and Ergonomics Society.
- Bourbousson, J., Poizat, G., Saury, J. and Sève, C. (2011). Description of dynamic shared knowledge: An exploratory study during a competitive team sports interaction. *Ergonomics*, 54(2), 120–138.
- Chase, W. G. and Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual Information Processing*. New York: Academic Press.
- Chi, M. T., Feltovich, P. J. and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices\*. *Cognitive Science*, 5(2), 121–152.
- Chiappe, D., Rorie, R. C., Morgan, C. A. and Vu, K. P. L. (2014). A situated approach to the acquisition of shared SA in team contexts. *Theoretical Issues in Ergonomics Science*, 15(1), 69–87.
- Crichton, M. T. and Flin, R. (2004). Identifying and training non-technical skills of nuclear emergency response teams. *Annals of Nuclear Energy*, 31(12), 1317–1330.
- Dekker, S. and Lutzhoft, M. (2004). Correspondence, cognition and sensemaking: A radical empiricist view of situation awareness. In *A Cognitive Approach to Situation Awareness: Theory and Application*. Burlington, VT: Ashgate Publishing Company (pp. 22–41).
- Dekker, S. W. (2000). Crew situation awareness in high-tech settings: Tactics for research into an ill-defined phenomenon. *Transportation Human Factors*, 2(1), 49–62.
- Dekker, S. W. (2013). On the epistemology and ethics of communicating a Cartesian consciousness. *Safety Science*, 56, 96–99.
- Dekker, S. W. and Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work*, 6(2), 79–86.
- Durso, F. T., Bleckley, M. K. and Dattel, A. R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(4), 721–733.
- Durso, F. T. and Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury and S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application* (pp. 137–154).
- Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, J. M. and Manning, C. A. (1998). En route operational errors and situational awareness. *The International Journal of Aviation Psychology*, 8(2), 177–194.
- Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *The International Journal of Aviation Psychology*, 3(2), 157–168.
- Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 32–64.
- Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 65–84.
- Endsley, M. R. and Garland, D. J. (Eds.). (2000). *Situation Awareness Analysis and Measurement*. Mahwah, NJ: Erlbaum.

- Endsley, M. R. and Rodgers, M. D. (October 1994). Situation awareness information requirements analysis for en route air traffic control. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 38, No. 1, pp. 71–75). Santa Monica, CA: Human Factors and Ergonomics Society.
- Fioratou, E., Flin, R., Glavin, R. and Patey, R. (2010). Beyond monitoring: Distributed situation awareness in anaesthesia. *British Journal of Anaesthesia*, 105(1), 83–90.
- Flach, J. M. (1995). Situation awareness: Proceed with caution. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 149–157.
- Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N. and Patey, R. (2003). Anaesthetists' Non-Technical Skills (ANTS): Evaluation of a behavioural marker system. *British Journal of Anaesthesia*, 90(5), 580–588.
- Fracker, M. L. and Vidulich, M. A. (1991). Measurement of situation awareness: A brief review. In Y. Queinnec and F. Daniellou (Eds.), *Designing for Everyone, Proceedings of the 11th Congress of the International Ergonomics Association* (pp. 795–797). London, UK: Taylor & Francis Group.
- Garbis, C. and Artman, H. (2004). Team situation awareness as communicative practices. In *A Cognitive Approach to Situation Awareness: Theory and Application*. Aldershot, U.K.: Ashgate.
- Gilson, R. D. (1995). Special issue preface. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 3–4.
- Golightly, D., Balfe, N., Sharples, S. and Lowe, E. (3–5 March 2009). Measuring situation awareness in rail signalling. In *Third International Conference on Rail Human Factors*, Lille, France.
- Golightly, D., Ryan, B., Dadashi, N., Pickup, L. and Wilson, J. R. (2013). Use of scenarios and function analyses to understand the impact of situation awareness on safe and effective work on rail tracks. *Safety Science*, 56, 52–62.
- Golightly, D., Wilson, J. R., Lowe, E. and Sharples, S. (2010). The role of situation awareness for understanding signalling and control in rail operations. *Theoretical Issues in Ergonomics Science*, 11(1–2), 84–98.
- Golightly, D., Wilson, J. R., Sharples, S. and Lowe, E. (2012). Developing a method for measuring situation awareness in rail signalling. In D. de Waard, N. Merat, A.H. Jamson, Y. Barnard and O.M.J. Carsten (Eds.), *Human Factors of Systems and Technology*. Maastricht, the Netherlands: Shaker Publishing.
- Gorman, J. C., Cooke, N. J. and Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, 49(12–13), 1312–1325.
- Gronlund, S. D., Ohrt, D. D., Dougherty, M. R., Perry, J. L. and Manning, C. A. (1998). Role of memory in air traffic control. *Journal of Experimental Psychology: Applied*, 4(3), 263.
- Hallbert, B. P. (June 1997). Situation awareness and operator performance: Results from simulator-based studies. In 'Global Perspectives of Human Factors in Power Generation', *Proceedings of the 1997 IEEE Sixth Conference on Human Factors and Power Plants* (pp. 18–1). IEEE.
- Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183.
- Hauss, Y. and Eyferth, K. (2003). Securing future ATM-concepts' safety by measuring situation awareness in ATC. *Aerospace Science and Technology*, 7(6), 417–442.
- Hogg, D. N., Folleso, K., Strand-Volden, F. and Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics*, 38(11), 2394–2413.
- Houghton, R. J., Baber, C., Cowton, M., Walker, G. H. and Stanton, N. A. (2008). WESTT (workload, error, situational awareness, time and teamwork): An analytical prototyping system for command and control. *Cognition, Technology & Work*, 10(3), 199–207.
- Hrebec, D. G. and Stiber, M. (April 2001). A survey of system administrator mental models and situation awareness. In *Proceedings of the 2001 ACM SIGCPR Conference on Computer Personnel Research* (pp. 166–172). New York: ACM Press.
- Jeannot, E., Kelly, C. and Thompson, D. (2003). The development of situation awareness measures in ATM systems. Brussels, Belgium: Eurocontrol. <http://www.skybrary.aero/bookshelf/books/223.pdf>. (Accessed 26th November 2014).
- Jones, D. G. (2000). Subjective measures of situation awareness. In M. R. Endsley and D. J. Garland (Eds.), (2000), *Situation Awareness Analysis and Measurement* (pp. 113–128). Mahwah, NJ: Erlbaum.
- Jones, D. and Endsley, M. (1996). Sources of situation awareness errors in aviation. *Aviation, Space and Environmental Medicine*, 67(6), 507–512.
- Jones, D. G. and Endsley, M. R. (July 2000). Examining the validity of real-time probes as a metric of situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 1, pp. 278–278). Santa Monica, CA: Human Factors and Ergonomics Society.



- Just, M. A. and Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kaber, D. B. and Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2), 113–153.
- Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K. and Prinzel III, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, 36(5), 447–462.
- Klein, G. A., Calderwood, R. and Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, 19(3), 462–472.
- Langan-Fox, J., Code, S. and Langfield-Smith, K. (2000). Team mental models: Techniques, methods, and analytic approaches. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(2), 242–271.
- Ma, R. and Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics*, 35(10), 939–953.
- Matthews, M. D., Martinez, S. G., Eid, J., Johnsen, B. H. and Boe, O. C. (September 2005). A comparison of observer and incumbent ratings of situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 3, pp. 548–551). Santa Monica, CA: Human Factors and Ergonomics Society.
- Matthews, M. D., Pleban, R. J., Endsley, M. R. and Strater, L. D. (October 2000). Measures of infantry situation awareness for a virtual MOUT environment. Measures of infantry situation awareness for a virtual MOUT environment. In: M. D. Matthews, R. J. Pleban, M. R. Endsley, & L. G. Strater (2000). *Proceedings of the Human Performance, Situation Awareness and Automation: User Centred Design for the New Millennium Conference* (pp. 262–267). Savannah, GA: SA Technologies.
- Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology*. Freeman, San Francisco.
- Paige Bacon, L. and Strybel, T. Z. (2013). Assessment of the validity and intrusiveness of online-probe questions for situation awareness in a simulated air-traffic-management task with student air-traffic controllers. *Safety Science*, 56, 89–95.
- Parasuraman, R., Sheridan, T. B. and Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2(2), 140–160.
- Rodgers, M. D., Mogford, R. H. and Strauch, B. (2000). Post hoc assessment of situation awareness in air traffic control incidents and major aircraft accidents. In: M. R. Endsley and D. J. Garland (Eds.). (2000). *Situation Awareness Analysis and Measurement* (pp. 73–112). Mahwah, NJ: Erlbaum.
- Salmon, P. M., Lenné, M. G., Young, K. L. and Walker, G. H. (2013). An on-road network analysis-based approach to studying driver situation awareness at rail level crossings. *Accident Analysis & Prevention*, 58, 195–205.
- Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Baber, C. and McMaster, R. (2008). Representing situation awareness in collaborative systems: A case study in the energy distribution domain. *Ergonomics*, 51(3), 367–384.
- Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L. and Young, M. (2009). Measuring situation awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39(3), 490–500.
- Sarter, N. B. and Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology*, 1(1), 45–57.
- Saus, E. R., Johnsen, B. H. and Eid, J. (October 2006). The relationship between dimensions of personality and situational awareness in a navigation simulator. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 17, pp. 1726–1729). Santa Monica, CA: Human Factors and Ergonomics Society.
- Saus, E. R., Johnsen, B. H., Eid, J., and Thayer, J. F. (2012). Who benefits from simulator training: Personality and heart rate variability in relation to situation awareness during navigation training. *Computers in Human Behavior*, 28(4), 1262–1268.
- Smith, K. and Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 137–148.
- Smolensky, M. W. (1993). Toward the physiological measurement of situation awareness: The case for eye movement measurements. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.



- Sneddon, A., Mearns, K. and Flin, R. (2006). Situation awareness and safety in offshore drill crews. *Cognition, Technology & Work*, 8(4), 255–267.
- Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C. and Jenkins, D. (2005). *Human Factors Methods: A Practical Guide for Engineering and Design*. Aldershot, U.K.: Ashgate.
- Stanton, N. A., Salmon, P. M., Walker, G. H. and Jenkins, D. (2009). Genotype and phenotype schemata and their role in distributed situation awareness in collaborative systems. *Theoretical Issues in Ergonomics Science*, 10(1), 43–68.
- Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R. and Green, D. (2006). Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology. *Ergonomics*, 49(12–13), 1288–1311.
- Strybel, T. Z., Vu, K.-P. L., Bacon, P. L., Billinghamurst, S., Rorie, R. C., Kraut, J. M., Morgan, C., Battiste, V. and Johnson, W. (2011). Situation awareness, workload, and performance in midterm NextGen: Effect of dynamic variations in aircraft equipage levels. *Proceedings of the 2011 International Symposium on Aviation Psychology* (pp. 118–123). Dayton, OH.
- Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a toll for aircrew systems design. In: *AGARD Conference Proceedings No 478, Situational Awareness in Aerospace operation*. (SEE N 90-28972 23-53).
- Taylor, R. M. and Selcon, S. J. (October 1990). Cognitive quality and situational awareness with advanced aircraft attitude displays. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 34, No. 1, pp. 26–30). Santa Monica, CA: Human Factors and Ergonomics Society.
- Tschirner, S., Andersson, A. W. and Bengt, S. (2013). Designing train driver advisory systems for Situation Awareness. In: N. Dadashi, A. Scott, J. Wilson and A. Mills (Eds.). *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction* (p. 150). London UK.: Taylor and Francis.
- Underwood, G., Chapman, P., Bowden, K. and Crundall, D. (2002). Visual search while driving: Skill and awareness during inspection of the scene. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 87–97.
- van Winsen, R., Henriqson, E., Schuler, B. and Dekker, S. W. (2014). Situation awareness: Some conditions of possibility. *Theoretical Issues in Ergonomics Science*, 16(1), 53–68.
- Vicente, K. J. (1999). *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. Mahwah, NJ: Erlbaum.
- Vidulich, M. A. and Hughes, E. R. (September 1991). Testing a subjective metric of situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 35, No. 18, pp. 1307–1311). Santa Monica, CA: Human Factors and Ergonomics Society.
- Vu, K. L., Strybel, T. Z., Kraut, J., Bacon, P., Minakata, K., Nguyen, J. and Johnson, W. (October 2010). Pilot and controller workload and situation awareness with three traffic management concepts. In *2010 IEEE/AIAA 29th Digital Avionics Systems Conference (DASC)* (p. 4-A). IEEE.
- Walker, G. H., Gibson, H., Stanton, N. A., Baber, C., Salmon, P. and Green, D. (2006). Event analysis of systemic teamwork (EAST): a novel integration of ergonomics methods to analyse C4i activity. *Ergonomics*, 49(12–13), 1345–1369.
- Wilson, G. F. (2000). Strategies for psychophysiological assessment of situation awareness. In: M. R. Endsley and D. J. Garland (Eds.). (2000). *Situation Awareness Analysis and Measurement* (pp. 175–188). Mahwah, NJ: Erlbaum.



---

# 20 Assessment of Potentially Stressful Work and Organisational Environments

## *Theory and Method*

*Tom Cox and Amanda Griffiths*

### CONTENTS

|                                                                                                 |     |
|-------------------------------------------------------------------------------------------------|-----|
| Introduction.....                                                                               | 565 |
| Work-Related Stress.....                                                                        | 566 |
| Psychosocial Hazards.....                                                                       | 567 |
| Reliability of Assessments .....                                                                | 568 |
| Approach.....                                                                                   | 568 |
| Assessment of Work and Organisational Environments .....                                        | 569 |
| Who Is an Expert?.....                                                                          | 570 |
| Reasoning and Measurement .....                                                                 | 570 |
| Taxonomy of Psychosocial Hazards .....                                                          | 571 |
| Psychosocial Taxonomy Based on Work and Organisational Characteristics .....                    | 571 |
| Psychosocial Taxonomy Based on Theoretical Constructs Relating to the Experience of Stress..... | 572 |
| Combining the Two Types of Taxonomy .....                                                       | 573 |
| Measuring the Experience of Stress.....                                                         | 574 |
| Stress Arousal Checklist.....                                                                   | 574 |
| Measuring Health at Work .....                                                                  | 575 |
| Theory of Health and Well-Being .....                                                           | 575 |
| Suboptimal Health.....                                                                          | 575 |
| General Well-Being Questionnaire .....                                                          | 576 |
| Final Commentary .....                                                                          | 578 |
| References.....                                                                                 | 579 |

### INTRODUCTION

In a civilised society, it is the right and duty of every man and woman to work in order to earn an adequate living to support themselves and their families and ensure their independence from the State. It is the duty of the State to provide appropriate opportunities for them to do so, to ensure that such work is without detriment to their health or risk to their safety and to support those who, though no fault of their own, cannot work. That these things are not always possible or do not otherwise occur is both a failing of and a challenge to our society. Preventing work-related stress and associated ill health is a part of that challenge.

Surveys of working people in Europe, North America and elsewhere have identified *work-related stress* as a major challenge to health and safety for working people. It has also been linked both to absence from work and to ill-health retirement. In England and Wales, stress has consistently been

one of the most commonly reported types of work-related illness cited in the national Labour Force Survey (LFS) (see Hodgson et al., 1993; Jones et al., 2003). A broadly consistent picture is painted by other sources of relevant data in England and Wales (see, e.g. Cherry et al., 2006; Carder et al., 2009). Based on the LFS data, the estimated cases of work-related stress, both prevalence and incidence, have remained high but stable over the past decade: 2000/2001 to 2010/2011. In 2011/2012, the prevalence of stress was 428,000 cases out of a total of 1,073,000 cases for all work-related illnesses (40%). The employment sectors that reported the highest rates of total cases of work-related stress (3 year average) were human health and social work, education and public administration and defence. The occupations that reported the highest prevalence rates (3 year averages) were health professionals (in particular, nurses), teaching and educational professionals and caring personal services (in particular, welfare and housing associate professionals). These occupations, taken overall, share a high level of interaction with the public which was suggested as a risk factor early in the contemporary history of work-related stress (Colligan et al., 1977). However, the main work activities attributed by respondents in the LFS as causing their work-related stress, or making it worse, were work pressure, lack of managerial support and work-related violence and bullying.

In Europe, a major step forward in meeting the challenge of work-related stress was taken in 1989 by the European Commission when it published its Framework Directive on the Introduction of Measures to Encourage Improvements in the Safety and Health of Workers at Work (EC, 1989). All Member States of the European Union were required then to incorporate these measures into their national legislation. The Directive required employers to avoid risks to employees' safety and health, to evaluate those risks that could not be avoided, to combat such risks at source (Article 6: 2) and to consult workers or their representatives and allow them to take part in discussions on all questions relating to safety and health at work (Article 11: 1). Employers were charged with developing a coherent overall prevention policy which covered technology, the organisation of work, working conditions, social relationships and the influence of factors related to the working environment (Article 6: 2). They had to make an assessment of the risks to safety and health at work and to decide on the protective measures to be taken (Article 9: 1). It was also stated in the 1989 Directive that risk assessment should involve a systematic examination of all aspects of the work undertaken to consider what could cause injury or harm, whether the hazards could be eliminated and, if not, what preventive or protective measures are, or should be, in place to control those risks (European Commission, 1996).

The 1989 Framework Directive, although more than 20 years old and now much developed in particular respects, provides the basis for this chapter and, in Europe, for the assessment of potentially stressful work and organisational environments.

This chapter discusses the nature of work-related stress and the assessment of potentially stressful work and organisational environments. It then considers how the reliability of such assessments can be strengthened through the triangulation of relevant data. In doing so, it outlines an approach – the *Cox–Griffiths methodology* – which triangulates the expert assessment of psychosocial hazards at work with the measurement of employees' experience of stress and measures of their health. It presents three instruments which have been used as part of this approach. The issues involved here are not straightforward given both the nature of the phenomenon and the complexity inherent in undertaking scientifically acceptable assessments and subsequent interventions in the context of real working life (Griffiths, 1999; Cox et al., 2007).

## WORK-RELATED STRESS

The history of the stress concept has been dealt with extensively in the literature and there is much written on contemporary theories of stress in relation to work (see, e.g. Cox, 1978; Cooper, 1998; Cooper and Dewe, 2004; Cox and Griffiths, 2010). Consistent with much of the current literature, this chapter adopts what has been termed an *interactional* or *transactional* approach to the definition of work-related stress. Stress is treated as a negative psychological state with significant cognitive and emotional components. The experience of work-related stress is held to arise through the

person's perception of the nature of the interaction between them and their work and organisational environments. Essentially, stressful work and organisational environments are those in which the demands on the person do not match their abilities, do not fulfil their needs and do not offer them sufficient control over work or sufficient support in coping with work (Cox, 1978). The experience of stress is inherent in the interaction between the person and their work. However, because the person is not passive in this interaction, but is an active agent, some theories refer to a transactional relationship (see, e.g. Cox and Griffiths, 2010; Dewe et al., 2012). The author's transactional approach to work-related stress provides the theoretical framework for the methodology described here. It is discussed in detail later in this chapter.

There is much evidential support for interactional and transactional theories of work-related stress (Cox et al., 2000). Within this genre of theory, there are many different variants, developed by different researchers or for different work groups, situations and organisations. Many of these differences are either semantic or ones of perspective, and there is a reasonably strong consistency in thinking at the more macro level. The variations in theory which exist account for the available data in slightly different ways but to more or less the same effect. A more detailed discussion of theories of work-related stress can be found in Cooper et al. (2001) and Cox and Griffiths (2010).

Work-related stress, defined as a psychological state, is both part of and reflects a wider process of interaction between people and their work and organisational environments. Part of that process is the sequence of relationships between the objective work and organisational environments and the worker's perceptions, between those perceptions and the emotional experience of stress and between that experience and associated changes in behaviour, physiological function and health. This sequence provides a basis for measurement in which individual perceptions and emotions are central. Logically, therefore, the measurement of work-related stress should be based primarily on self-report measures which focus on the appraisal process or on the emotional experience of stress (Cox, 1985).

Demands are not inherently stressful. They are experienced as stressful when they do not match the person's abilities and when the person does not have adequate control or support in coping with them. The work and organisational characteristics which are experienced as stressful may be treated as hazardous to health and may affect psychological and physical health as well as emotional well-being. The evidence suggests that the emotional experience of stress associated with work may arise in one of two ways: first, through anxiety over exposure to the more traditional and tangible hazards of work and, second, through exposure to the psychosocial hazards of work (Cox, 1993; Cox et al., 2000). Exposure to the latter is the focus of attention here.

## PSYCHOSOCIAL HAZARDS

Arguably our concern for the effects of work on health and safety can be traced back to the Middle Ages in Europe (Schilling, 1981). Examples exist of the actions taken in the then developing mining industry not only to protect miners through advances in rudimentary mine engineering and ventilation but also through the use of basic personal protective equipment. Action was also taken, at the community level, to treat miners who were ill or injured through work and to support their families in the face of their disability and death. Much is now known about the nature and mechanisms of effect of these more traditionally conceived and more tangible hazards. Our ability to intervene to prevent and manage associated injury, illness and death has advanced greatly over the last 100 years. However, with significant changes in the nature of work and work organisations, working life and work values (see, e.g. Quam, 2010; Wood, 2011; Kalleberg and Marsden, 2012), new hazards to employee health have emerged. Over the last 40 or so years, recognition has had to be given to a new set of challenges, those inherent in failures in the design and management of work and of work organisations (Cox, 1998; European Agency for Safety and Health at Work, 2010). These hazards have been economically termed 'psychosocial hazards'.

The International Labour Organization (ILO) was one of the first bodies to use the term *psychosocial hazard*. In a joint report with the World Health Organization (WHO) in the mid-1980s

(ILO and WHO, 1986), it summarised the evidence that suggested that the way in which work was structured and managed and the way in which work organisations ran affected the health of employees as well as their behaviour and productivity at work. It termed failures of work design and management, and those of the work organisation, psychosocial hazards. The report also suggested that the mechanism underpinning the harm caused by psychosocial hazards was related to the experience of stress through work. From these beginnings, a new area of health and safety expertise has developed around psychosocial risk assessment and management (Leka and Cox, 2008; Leka et al., 2008; Leka and Jain, 2010; Cox, 2011). The assessment of potentially stressful work and organisational environments, focused on exposure to psychosocial hazards, is central to this area.

## RELIABILITY OF ASSESSMENTS

It is well established that, among other things, all assessments must be reliable, valid and fair. Among these various characteristics, arguably, reliability is first among equals. It is meaningless to discuss the validity and fairness of instruments which cannot offer reliable measurement. Reliability can be assured for quantitative data in, at least, two different ways: first, through their proper psychometric development involving statistical proof of reliability and, second, through the comparison of data acquired using different methods and the establishment of substantive consensus (see Chapter 1 for further discussion of reliability and validity). For qualitative data, reliance is largely based on the latter. Given that the psychometric instruments discussed later in this chapter have been properly developed, the emphasis here is on the comparison of data using different methods of assessment.

It takes, at least, three points on a graph to establish a straight line reliably or topographically to establish a location reliably. Extending this principle, it has been argued that the comparison of a minimum of three different types of data is required to establish the reliability of an assessment instrument. This is termed the *triangulation\** of data (see, e.g. Denzin, 2006, 2012).

How *different* the various types of data must be is a mute point. There are at least three issues to be considered here in the assessment of work and organisational environments: what is being measured and its appropriateness (validity), the method of measurement and how practical is the use of the measure in real-world situations. Ideally, the measures used in triangulation, while all being appropriate and practical, should differ in their focus (what is being measured) and their methodology (how). However, a trade-off may be justified between differences in focus and in methodology. For example, the same aspect of a situation might be assessed using two different methods or different but appropriate aspects might be assessed using the same method. In the approach suggested here, the latter is the case. The critical point for triangulation is that the measurements of the different types of data used are independent.

## APPROACH

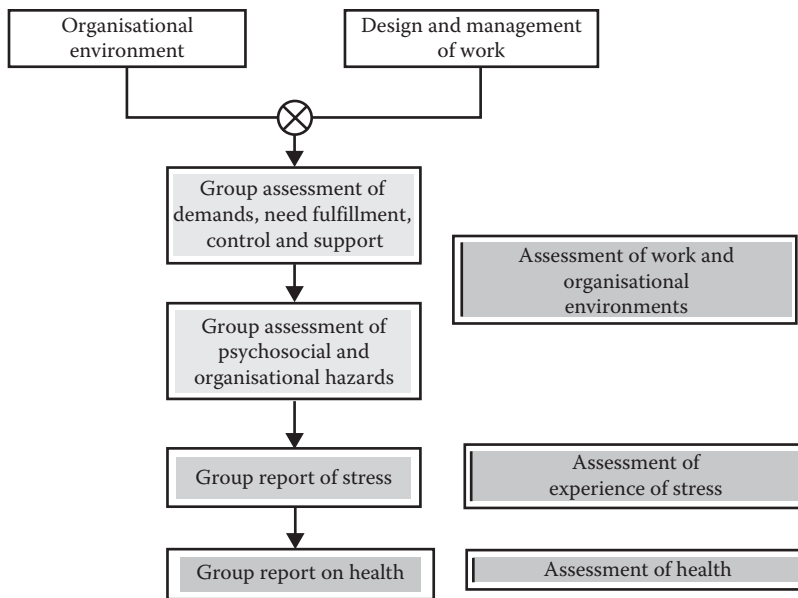
The approach, suggested here, has been referred to as the *Cox–Griffiths methodology*. It is based on the triangulation of three independent assessments of the correlates of work-related stress which together capture the notion of an interaction between the person and their work. The three assessments are

1. An expert assessment of the work and organisational environments
2. The psychometric assessment of the experience of stress reported by those working in such conditions
3. The psychometric assessment of their health

---

\* In many areas of research method discussion, including this book, the term triangulation has in fact been used to refer to the application of *two* or more methods (see Chapters 1 and 5); in this chapter we use the term in its *true* sense derived from its geospatial origins.





**FIGURE 20.1** Cox–Griffiths methodology.

Together these three assessments can offer a description of the process underpinning the nature and effects of work-related stress on employees' health (see Figure 20.1): environmental antecedents of stress → the experience of stress → associated ill health. Somewhat similar approaches can be found in the literature. For example, Bailey and Bhagat (1987) recommended a multi-method approach to the measurement of stress consistent with triangulation. They argued in favour of balancing the evidence from the self-report of work and the experience of work with physiological and unobtrusive measures. Unobtrusive measures relate to what Folger and Belew (1985) called non-reactive measures and include physical traces (such as poor housekeeping), archival data (such as absenteeism records), private records (such as diaries) and (unobtrusive) observation.

It is important to note two important caveats in relation to the *Cox–Griffiths methodology*. First, the approach has been developed for the assessment of potentially stressful work and organisational environments. It should be used with a focus on defined types of work and organisational situations and with defined populations or population samples. It has not been designed for use with individual cases. Second, the purpose of the approach is to offer a *snapshot* in time with reference to appropriate normative data or for comparison in situ between appropriate groups. Like many instruments used in this domain, it has proved useful in the measurement of change, possibly following intervention, but the assessment of change per se is not built into the method. The exploration of change in the assessments has to be managed through either experimental or adequate quasi-experimental design (see, e.g. Thyer, 2012).

## ASSESSMENT OF WORK AND ORGANISATIONAL ENVIRONMENTS

The first of the three sets of data that have been used as part of the *Cox–Griffiths methodology* is based on the expert assessment of the work and organisational environments. This requires three things: agreement on and the availability of appropriate experts, an evidence-based taxonomy of the psychosocial hazards to be assessed and the nature of the reasoning used in the assessment.

## WHO IS AN EXPERT?

The basis for expertise is knowledge supported by appropriate experience (Argote and Miron-Spektor, 2011). In this context, knowledge might be developed through education or research, through reflections on practical experience (in work) or, optimally, through a combination of both. Organisations tend to be quick to recognise expertise that involves education and research. However, in the present context, they have been somewhat slower to accept that which comes through actual on-the-job experience: doing the work. Organisations appear more likely to *trust* the educated, qualified and often external expert than to involve their own employees, those who actually carry out the work in question on a day-to-day basis. The issues are not straightforward, but the failure to involve employees, who *own* the work in question through their day-to-day engagement with it, could be a mistake (Mellor et al., 2013). Furthermore, for those based in the European Union, it is a requirement under the 1989 European Directive to ensure employee involvement in risk assessments or that of their representatives.

The *Cox–Griffiths methodology* recommends such employee involvement for, at least, three reasons: logical, moral and instrumental. First, it *ensures a logical consistency* within the method (using employee-generated data) and between the method and its purpose (improving employees' work and organisational environments). Second, it *properly involves those who own the problem of work-related stress* in the assessment of those environments and, third, it most likely *starts to build their engagement in and commitment to any later intervention*.

The involvement of employees in the assessment of the potential stressfulness of their work and organisational environments is part of the *Cox–Griffiths methodology*. However, it is not meant to exclude the use of other types of expert. This is especially so where there is an obvious need for a specific type of expertise to drill down into particular risk factors which the method has identified as recommended by the GB Health and Safety Executive. For example, if the application of this method identifies problems with workstation design, then a suitably qualified ergonomist might be consulted to explore and correct those problems. The medical analogy is that the doctor in primary care may refer a patient to an appropriate specialist once he or she has identified the likelihood of a particular and difficult condition. All of this requires that those employees involved understand the nature and purpose of the exercise and have been briefed on the methods that they will be asked to use. In the authors' experience, this has never proved an obstacle being neither time nor resource hungry.

The question of the use of employees as experts in the assessment of their work and organisational environments, in turn, raises the question of how they should make such an assessment: the type of reasoning that they should use.

## REASONING AND MEASUREMENT

Consistent with Locke and Pennington (1982), there appear to be, at least, two different types of scale that can be used in the assessment of work and organisational environments. They are based on two different types of reasoning. The first, and most common, is essentially about the individual's reactions to their work and organisational environments. This is *psychological reasoning* and by-and-large attempts to capture answers to questions about how the individual feels about his or her work and organisational environments or how these *affect their mood or emotional state*. These questions about feelings, mood and well-being are, by necessity, context dependent. The second, and in the present context more appropriate type of reasoning, is focused on the situation and not on the individual. This is *situational reasoning*. It asks the respondents to use their experience and knowledge to judge the adequacy or problematic nature of different aspects of the work and organisational environments and in terms of (all) those who are involved. The data collected from the defined population or population sample under study can be aggregated, as necessary, and the group response taken and used.

## TAXONOMY OF PSYCHOSOCIAL HAZARDS

To be reliable and valid, any assessment has to use a standard set of questions and a proven theoretical or empirical framework. The framework suggested for the *Cox–Griffiths methodology* is based on existing taxonomies of psychosocial hazards. There have been two different approaches to the taxonomy of such hazards, and this is reflected in the nature of the research carried out in this area and in the types of interventions attempted. They differ in their granularity of measurement.

The first and more detailed approach is focused on the level of particular work and organisational characteristics. Its use is often driven by a concern to collect sufficient information about work to allow the identification of actual problems and to inform intervention. The second approach, at a higher conceptual level, considers the dimensions underpinning psychosocial hazards in general and is usually in terms of one of the dominant theories of work-related stress (Cox and Griffiths, 2010). It is more theoretically driven with the objective of better understanding the nature of psychosocial hazards. Either approach might be used in the assessment of work and organisational environments. The two approaches can be combined.

### PSYCHOSOCIAL TAXONOMY BASED ON WORK AND ORGANISATIONAL CHARACTERISTICS

The taxonomy of psychosocial hazards suggested here was first presented by Cox in the early 1990s (Cox, 1993; Cox and Cox, 1993) and has been updated on several occasions since then (Cox et al., 2000; Cox, 2011). All have been based on reviews of the available literature (e.g. Kivimaki et al., 1997; Stansfeld and Candy, 2006; Nieuwenhuijsen et al., 2010) and modified by the results of empirical studies. Each taxonomy was put forward as an easily applicable basis for a crude assessment which would act as the starting point for a process of intervention and improvement. The basic taxonomy is summarised in Table 20.1.

In many of the studies used in the development of the *Cox–Griffiths methodology*, a simple five-point scale of adequacy or problematic nature was used. To avoid leading the respondent's judgements, where possible the scales were bipolar with a neutral midpoint. In addition, a *not applicable* response category was used. The scales were worded in a way appropriate to the population or population sample under study and were piloted. The instrument used in these studies has been termed the 'Work Environment Survey'.

There are a number of points to note about this taxonomy.

First, the taxonomy refers to work and organisational characteristics which, if they fail in some sense or are inadequate or lacking, have the potential to cause harm to employees' health. These characteristics are the psychosocial hazards. However, many also have the potential, if optimal, to benefit employees. In this respect, psychosocial hazards are unlike many traditionally conceived and more tangible hazards. Second, a distinction can be made, following Hacker (1978), between those psychosocial hazards that relate to the nature of work itself and are task or job related and those that relate to the organisational context to work. This hierarchical way of thinking about psychosocial hazards could be extended further to include those work-related challenges to employees' health which originate in working life and the wider community or which reflect the more general economic and political landscape. Third, workplace violence and bullying are included in this taxonomy under the more general heading: interpersonal relationships at work. However, fatigue has not been included in this taxonomy as it is an outcome in the underlying model rather than a cause. Furthermore, change, per se, is not included in this taxonomy as it underpins or exacerbates much of what is potentially harmful in relation to the work and organisational characteristics listed.

It was never claimed that this taxonomy was exhaustive or, in any sense, final given the changing work landscape and a relatively recent publication by the European Agency on Safety and Health at Work (2010) has summarised expert opinion on likely new risks to health and safety in the coming years.

**TABLE 20.1**  
**Psychosocial and Organisational Hazards**

| Psychosocial Hazard                 | Conditions                                                                                                                                                              |
|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Content of work</i>              |                                                                                                                                                                         |
| Job content                         | Lack of variety or short work cycles, fragmented or meaningless work, underuse of skills, high uncertainty, continuous exposure to people through work                  |
| Workload and work pace              | Work overload or underload, machine pacing, high levels of time pressure, continually subject to deadlines                                                              |
| Work schedule                       | Shift working, night shifts, inflexible work schedules, unpredictable hours and long or unsociable hours                                                                |
| Control                             | Low participation in decision-making, lack of control over workload, pacing, shift working, etc.                                                                        |
| Physical environment and equipment  | Inadequate equipment availability, suitability or maintenance; poor environmental conditions such as lack of space, poor lighting, excessive noise                      |
| <i>Context to work</i>              |                                                                                                                                                                         |
| Organisational culture and function | Poor communication, low levels of support for problem-solving and personal development, lack of definition of, or agreement on, organisational objectives and practices |
| Interpersonal relationships at work | Social or physical isolation, poor relationships with superiors, interpersonal conflict, violence and bullying, lack of social support                                  |
| Role in organisation                | Role ambiguity, role conflict and responsibility for people                                                                                                             |
| Career development                  | Career stagnation and uncertainty, underpromotion or overpromotion, poor pay, job insecurity, low social value to work                                                  |
| Home–work interface                 | Conflicting demands of work and home, low support at home, dual-career problems                                                                                         |

*Sources:* Adapted from Cox, T., *Stress Research and Stress Management: Putting Theory to Work*, HSE Books, Sudbury, 1993; Cox, T. et al., *Work Stress*, Office for Official Publications of the European Communities, Luxembourg, 2000.

## PSYCHOSOCIAL TAXONOMY BASED ON THEORETICAL CONSTRUCTS RELATING TO THE EXPERIENCE OF STRESS

There are four key constructs which define the *transactional approach* to work-related stress (Cox, 1978; Cox and Griffiths, 2010): demands, need fulfilment, control and support. These are judged in relation to the individual's abilities, coping resources and requirements. In terms of assessing the work and organisational environments, they have been presented as

1. The extent to which the demands made on the person are matched to their knowledge, abilities and skills encompassing underload as well as overload
2. The extent to which work and the organisation fulfil the person's needs including those arising from their beliefs and values
3. The extent to which the person has control over their work and how they cope with their work and organisational environments
4. The extent to which the support available to the person in coping with their work and organisational environments is adequate

In the reality of working life, each of the issues is *a matter of degree*. Five-point bipolar scales have been used by the authors with a mid-scale neutral point and with each scale point suitably worded for the population or population sample used. A *not applicable* response category has not been used.

In relation to the measurement of demand, Dewe (1991) has argued that it is necessary to go beyond simply asking whether particular demands are present in their work and organisational

environments. Because of the very nature of demand, he argues that there is a need to measure dimensions such as frequency, duration and level. It is argued here that any more detailed measures of demand need to be applied in a way that allows for the possibility of interaction, such as those between demand and control (Karasek, 1979; Warr, 1990) or demand and control with support (Cox, 1978, 1985; Karasek, 1979; Karasek and Theorell, 1990). Of course, the person's ability to deal with the demands of work needs also to be taken into account (as being discussed here).

## COMBINING THE TWO TYPES OF TAXONOMY

Logically, there are, at least, two ways to combine these different types of taxonomy (work and organisational and theoretical constructs), either by using them in parallel or by developing a hybrid from key elements of each. The former is relatively straightforward.

Arguably one of the more influential taxonomies of psychosocial hazards in the United Kingdom has been developed for the management standards approach to work-related stress by the Health and Safety Executive (HSE). It was described by Mackay et al. (2004) and offers an example of a hybrid taxonomy:

The taxonomy developed from Cox's research (Cox, 1993; Cox et al., 2000), in conjunction with the findings of other HSE studies (for example, Smith et al., 2000; Stansfeld et al., 2000) formed the basis for the approach adopted in the development of draft standards for the good management of work-related stress, which have been named Management Standards.

In the late 1990s, the HSE held a series of workshops at which the issue of a practical psychosocial taxonomy was discussed. It then reviewed its commissioned research on the subject and also all existing taxonomies and associated theories. As a result of this work, a group of seven hazard areas was agreed. These formed the basis of HSE guidance *Tackling Work-Related Stress* (HSG218) (HSE, 2001) as follows:

- *Demands* (including such issues as workload, work patterns and the working environment)
- *Control* (how much say the person has in the way they do their work)
- *Support* (which includes the encouragement, sponsorship and resources provided by the organisation, line management and colleagues)
- *Relationships at work* (which includes promoting positive working practices to avoid conflict and dealing with unacceptable behaviour)
- *Role* (whether people understand their role within the organisation and whether the organisation ensures that the person does not have conflicting roles)
- *Change* (how organisational change [large or small] is managed and communicated in the organisation)
- *Culture* (the way in which organisations demonstrate management commitment and have procedures which are fair and open)

The influence of interactional theories, especially that of Karasek (see Karasek, 1979; Karasek and Theorell, 1990) and of the Michigan School (see, e.g. Kahn et al., 1964), is obvious in the architecture of this taxonomy. The detail, nested within this architecture, is based on work characteristics commonly held to be hazardous to employees' health.

In the subsequent development of the Management Standards, culture as a separate topic was dropped because it underpins the approach to each of the others. Aspects of culture are incorporated into each of the remaining six. This is a similar argument to that advanced here concerning change. However, while it can be sustained for change, the loss of a culture issue here means that an important area of hazard might be easily overlooked. As with the Cox (1993) taxonomy, fatigue was not included.

## MEASURING THE EXPERIENCE OF STRESS

Following the arguments presented earlier in this chapter about the nature of stress, it would seem that the measurement of mood may offer one direct method of tapping the emotional experience of stress. In the late 1970s and early 1980s, there was a surge of interest in this issue as witnessed by a series of articles in the *British Journal of Psychology* (Cruickshank, 1982, 1984; King et al., 1983; Cox and Mackay, 1985) and elsewhere (Burrows et al., 1977; Russell, 1979, 1980; Ray and Fitzgibbon, 1981; Watts et al., 1983). Many of these studies employed the Stress Arousal Checklist (SACL) developed by Cox and Mackay and originally published in the *British Journal of Social and Clinical Psychology* (Mackay et al., 1978). There were also a number of studies which used modified versions of the checklist (Cruickshank, 1982, 1984) although, sometimes, locally inspired changes in the instrument could not be defended (Cox and Mackay, 1985).

Since then, there has been sustained interest in the subject. A search of the available literature since 1990 revealed several thousand refereed articles in good-quality journals using the SACL. These tended to be of two types. First, there are those studies which used the original instrument, collectively, in a wide variety of different settings. Examples of relatively recent publications in different areas include psychophysiological function (Stadler et al., 2010), work (Hart et al., 1995, 2000), health and health behaviour (Heaney et al., 1998; Evatt and Kassel, 2010; Yamaguchi and Kanemitsu, 2011) and education (Chung et al., 2012; Faleel et al., 2012). Second, there are those studies which have sought to develop the instrument further (see, e.g. McCormick et al., 1985; Matthews et al., 1990). The latter are useful in that all psychometric instruments must benefit from continuing development. However, the former stand testament to the fact that the original instrument still works well.

### STRESS AROUSAL CHECKLIST

The SACL is an adjective checklist and was developed using factor analytical techniques (Mackay et al., 1978; Cox and Mackay, 1985; Gotts and Cox, 1990) for the measurement of self-reported mood. It presents the respondent with 30 relatively common mood-describing adjectives and asks to what extent they describe their current feelings. The model of mood which underpins the checklist is two dimensional. One dimension appears to relate to feelings of unpleasantness/pleasantness or hedonic tone (stress) and the other to wakefulness/drowsiness or vigour (arousal). Such a model is well represented in the relevant psychological and psychophysiological literature (see, e.g. Mackay, 1980; Russell, 1980). The split half reliability coefficients for the two scales which tap into these dimensions have always proved acceptable: for example, arousal 0.82 and stress 0.80 (Watts et al., 1983; Gotts and Cox, 1990). Both were conceived of and developed as state measures and are thus seen as transient in nature. The statistical device of test-retest coefficients is therefore not appropriate as a test of reliability.

Consistent with the transactional model (see earlier text), it was suggested by Mackay et al. (1978) that the stress dimension may reflect the perceived favourability of the external environment and thus have a strong cognitive component in its determination. Arousal, as it was suggested, might relate to ongoing autonomic and somatic activity and be essentially psychophysiological in nature. It became obvious that stress may partly reflect how appropriate the level of arousal is for a given situation and the effort of compensating for inappropriate levels (Cox et al., 1982). Together the two dimensions can be used to describe a four-quadrant model of mood within which characteristic emotions and related states may be identified: high arousal and high stress (anxiety), high arousal and low stress (pleasant excitement), low arousal and high stress (boredom) and, finally, low arousal and low stress (relaxed drowsiness).

A third scale has been suggested based on the use of a ‘?’ category on the response scale associated with the different mood adjectives. This category signifies, in part, uncertainty about whether



**TABLE 20.2**  
**Some Normative Data for the SACL (Derived from Gotts and Cox, 1990<sup>a</sup>)**

| Sample                | Dichotomised Scores |           |          |          |           |          |          |           |          |
|-----------------------|---------------------|-----------|----------|----------|-----------|----------|----------|-----------|----------|
|                       | Stress              |           |          | Arousal  |           |          | Q        |           |          |
|                       | <i>x</i>            | <i>SD</i> | <i>n</i> | <i>x</i> | <i>SD</i> | <i>n</i> | <i>x</i> | <i>SD</i> | <i>n</i> |
| Mixed population      | 6.0                 | 4.6       | 1027     | 6.4      | 3.2       | 1040     | 4.2      | 4.2       | 1079     |
| Males: mixed sample   | 6.0                 | 4.7       | 296      | 6.6      | 3.2       | 297      | 4.9      | 4.6       | 266      |
| Females: mixed sample | 6.0                 | 4.6       | 731      | 6.3      | 3.3       | 743      | 3.9      | 4.1       | 584      |
| Students              | 6.3                 | 4.9       | 515      | 5.7      | 3.6       | 518      | 4.7      | 4.1       | 535      |
| Ages 16–30            | 6.2                 | 4.6       | 466      | 6.0      | 3.2       | 469      | 5.0      | 4.3       | 379      |
| Ages 31–45            | 5.9                 | 4.9       | 344      | 7.2      | 3.3       | 353      | 3.5      | 4.2       | 334      |
| Age more than 45      | 5.1                 | 4.2       | 122      | 6.4      | 3.3       | 123      | 3.7      | 4.1       | 1132     |

<sup>a</sup> More complete normative data have been published as part of a manual for the SACL (Gotts and Cox, 1990). Further information can be obtained from the authors.

the adjective given currently describes the respondent's mood. A score based on the frequency of '?' responses might reflect an inability to report feelings, and this may be symptomatic of a disordered psychophysiological state. Such a scale has an acceptable split half reliability coefficient: 0.89 (Cox and Mackay, 1985).

A compilation of the available British and Australian data has allowed the publication of mean levels for different groups, broken down by country of origin, age, sex and occupation (Gotts and Cox, 1990). Some of these *normative* data are presented in Table 20.2.

## MEASURING HEALTH AT WORK

The measurement of health at work has tended to focus on death (mortality), work-related disease, disability or injury (morbidity) or general malaise. Fortunately, it appears that the experience of work-related stress is more likely to be associated with changes in the level of general malaise than in the incidence of death, or of disease, disability or injury (see Cox et al., 2000).

## THEORY OF HEALTH AND WELL-BEING

Health has been defined as a changeable state along a continuum from complete healthiness to death (World Health Organization, 1946; Rogers, 1960). This broad definition implies that health cannot be equated with mere absence of obvious disease, injury or disability and is not restricted to the sound physical condition of the body; it also has psychological and social aspects. In this context, *well-being* relates to individuals' experience of their health.

An important watershed in the health continuum is represented by the point where disease, disability or injury becomes obvious to the person and is usually represented by objectively verifiable and clinically significant signs and symptoms. Some of these will be diagnostic of a particular condition; others will more generally reflect the impact of that condition. Some will inevitably represent the effects of stress experienced in relation to being ill.

## SUBOPTIMAL HEALTH

The zone between complete physical, psychological and social healthiness, on the one hand, and obvious disease, disability or injury, on the other hand, has been termed 'suboptimal health' (Rogers, 1960). It has been suggested that suboptimal health may be represented as an experiential pool of

signs and symptoms of general malaise, each – on its own – of no particular clinical significance and certainly not diagnostic of any particular condition. Such signs and symptoms may or may not be precursors to disease, injury or disability depending on the operation of a wide range of health risk and salutogenic factors. At any time, different groups of signs and symptoms within the experiential pool will imperfectly predict particular ill-health outcomes. As a condition develops, the predictive group will refine itself and attract new signs and symptoms, and the prediction itself may strengthen. As a condition weakens, or the person recovers, then the reverse process will occur. It is *suboptimum health* that, as a concept and experience, equates most closely to the popular understanding of *well-being* and the two terms are used more or less interchangeably here.

A person considered to be normally healthy, by themselves or others, or more particularly judged not to be ill, will still experience something by way of such signs and symptoms of general malaise, and the possible pool of such experiences will be formally present at the group level. At this level, structural modelling will reveal a pattern and clusters of signs and symptoms reflective of an underlying *normal experience model*. However, it has been suggested that the normal experience of well-being may both reflect the experience of stress as one mediator of the effects of life and working conditions and also in turn affect other responses to stress, such as self-reported mood (see Mackay et al., 1978; Cox and Mackay, 1985). The *General Well-Being Questionnaire (GWBQ)* was developed within the framework of this theory of health.

### GENERAL WELL-BEING QUESTIONNAIRE

It was in the mid-1980s that the first author began to build a measurement tool for well-being based on the self-report of signs and symptoms of general malaise (Cox et al., 1983, 1984; Cox and Brockley, 1984). There were several different questionnaire instruments available at the time that by the nature of their scales and internal structure offered some description of that area of health (Gurin et al., 1960; Crown and Crisp, 1966; Goldberg, 1972; Derogatis et al., 1974). However, none of these were judged to be exactly what was required for use with a more or less healthy working population in Britain and for the purpose of assessing work-related stress.

Initially a compilation of non-specific symptoms of general malaise was produced from existing health questionnaires (see earlier text) and from diagnostic texts. These symptoms included reportable aspects of cognitive, emotional, behavioural and physiological function, none of which were clinically significant in themselves. From this compilation, a prototype checklist was designed with each symptom being associated with a five-point frequency scale (*never* through to *always*) which referred to a 6-month response window.

In a series of classical factor analytical studies, on British participants, variously reported (Cox et al., 1983, 1984), two clusters of symptoms or factors were identified (see Table 20.3). These factors were derived as orthogonal. The first factor (GWF1) was defined by symptoms relating to tiredness, emotional lability and cognitive confusion; it was colloquially termed ‘worn out’. The more cognitive items would appear to imply difficulties in decision-making (in the specific context of feeling *worn out*): (1) Has your thinking got mixed up when you have had to do things quickly? (2) Has it been hard for you to make up your mind? (3) Have you been forgetful? These may have implications for personal problem-solving and coping (see Cox, 1987). The second factor (GWF2) was defined by symptoms relating to worry and fear, tension and physical signs of anxiety; it was colloquially termed ‘uptight (or anxious) and tense’. This model of suboptimum health appeared to have some face validity in that it was acceptable to a conference audience of British general practitioners and medical and psychological researchers (see Cox et al., 1983).

It is therefore suggested by the authors that suboptimum health, the *grey area* between complete healthiness and obvious illness, is made up of two states, one related to being *worn out* and the other related to being *uptight and tense*. The former has an interesting cognitive component, possibly related to decision-making and coping, while the latter is partly defined by physical symptoms of anxiety and tension. It has been shown that people vary in the extent to which they report these

**TABLE 20.3**  
**Items Defining the GWBQ Scales (International Version)**

GWF1

Have your feelings been hurt easily?  
 Have you got tired easily?  
 Have you become annoyed and irritated easily?  
 Have your thinking got mixed up when you have had to do things quickly?  
 Have you done things on impulse?  
 Have things tended to get on your nerves and wear you out?  
 Has it been hard for you to make up your mind?  
 Have you got bored easily?  
 Have you been forgetful?  
 Have you had to clear your throat?  
 Has your face got flushed?  
 Have you had difficulty in falling or staying asleep?

GWF2

Have you worn yourself out worrying about your health?  
 Have you been tense and jittery?  
 Have you been troubled by stammering?  
 Have you had pains in the heart or chest?  
 Have you unfamiliar people or places made you afraid?  
 Have you been scared when alone?  
 Have you been bothered by thumping of the heart?  
 Have people considered you to be a nervous person?  
 When you have been upset or excited has your skin broken out in a rash?  
 Have you shaken or trembled?  
 Have you experienced loss of sexual interest or pleasure?  
 Have you had numbness or tingling in your arms or legs?

feelings, both between individuals and across time, and it has been suggested that this variation may not only (1) reflect the experience of stress but also (2) affect other responses to stress such as self-reported mood (see Mackay et al., 1978; Cox and Mackay, 1985). There is evidence that in workplace studies the worn-out scale shows greater utility than does the tense and uptight scale; it demonstrates a more consistent relationship with other, non-health measures of interest within the workplace (Cox et al., 2000).

This research culminated in the publication in English of the GWBQ (Cox et al., 1983).

In the late 1980s, more data were collected through a series of linked studies in Britain and Australia. These data were re-analysed, and the model and its associated scales were amended to increase their robustness in relation to this international sample and also to a diversity of homogeneous samples (see Table 20.4).

A number of symptoms (items) were deleted from the original scales, but no new symptoms were added. The two new international scales were each defined by 12 symptoms but retained their essential nature: worn out and tense and uptight. The deleted symptoms were among the weaker ones in terms of scale definition and item loadings. The early questionnaire was revised, new norms were computed and an international version was published and has been in use since then (e.g. Cox and Griffiths, 1995; Cox et al., 2000).

The possibility of cultural-linguistic differences in the experience and report of health has been explored in working populations in Taiwan (Ruey-Fa, 1994) and in Singapore (Ho, 1996). For example, bilingual Taiwanese school teachers (English and Mandarin) completed English and Mandarin versions of the GWBQ. The data clearly showed the emergence of identical two-factor models from

**TABLE 20.4**  
**Some Normative Data for the GWBQ (Derived from**  
**Unpublished Data of Cox and Gotts<sup>a</sup>) for Mixed Populations**

| Sample                        | International Version (1987) |           |          |                       |           |          |
|-------------------------------|------------------------------|-----------|----------|-----------------------|-----------|----------|
|                               | Worn Out<br>(12 Items)       |           | <i>n</i> | Uptight<br>(12 Items) |           |          |
|                               | <i>x</i>                     | <i>SD</i> |          | <i>x</i>              | <i>SD</i> | <i>n</i> |
| All                           | 16.7                         | 8.3       | 2300     | 10.7                  | 7.4       | 2312     |
| Males                         | 15.9                         | 7.8       | 1031     | 8.2                   | 6.5       | 1042     |
| Females                       | 17.4                         | 8.6       | 1262     | 12.8                  | 7.4       | 1262     |
| British sample by age (years) |                              |           |          |                       |           |          |
| 16–20                         | 16.5                         | 8.7       | 141      | 11.5                  | 7.9       | 141      |
| 21–25                         | 16.9                         | 9.2       | 147      | 11.3                  | 7.6       | 147      |
| 26–30                         | 15.6                         | 8.4       | 236      | 10.2                  | 7.5       | 236      |
| 31–35                         | 17.2                         | 8.6       | 239      | 9.0                   | 6.5       | 239      |
| 36–40                         | 16.1                         | 8.1       | 201      | 9.2                   | 7.5       | 201      |
| 41–45                         | 15.5                         | 8.6       | 199      | 10.4                  | 7.7       | 199      |
| 46–50                         | 16.0                         | 8.3       | 175      | 9.7                   | 7.7       | 175      |
| 51–55                         | 14.5                         | 8.0       | 174      | 9.1                   | 7.4       | 174      |
| 56–60                         | 13.7                         | 8.0       | 127      | 7.7                   | 6.6       | 127      |
| >60                           | 13.5                         | 6.4       | 26       | 4.8                   | 5.8       | 26       |

<sup>a</sup> Further information can be obtained from the authors.

these data sets with teachers' scores on the two versions being highly correlated. The Taiwanese (Mandarin) model was indistinguishable from the UK English-language version. Broadly similar data were obtained in the development of a Spanish-language version of the work-out scale (Cox et al., 2006). The instrument has also been successfully translated into Japanese (Leka et al., 2012). Despite these tests, the cultural-linguistic inter-changeability of the GWBQ should not be taken for granted and is a matter of empirical test between its English origins and other cultural-linguistic situations.

The GWBQ has been used extensively in studies of health and well-being with *normal* populations and in a variety of different contexts (see, e.g. in personality studies [Singh and Woods, 2008]; with organisational change [McHugh, 1997]; in the manufacturing sector (Griffiths et al., 2006); in police work [Hart et al., 1995]; with Swedish carer of the elderly [Dackert, 2010]; with Chinese office workers [Houdmont et al., 2011]; with Malaysian rail commuters [Mohd Mahudin et al., 2012]; and with Japanese psychiatric nurses [Leka et al., 2012]).

## FINAL COMMENTARY

Work-related stress is an important contemporary issue in occupational health, and there is a growing literature concerned with its nature and its measurement. The current consensus defines stress as a psychological state, with cognitive and emotional components, and acknowledges its effects on employees' health. Furthermore, there is evidence that the experience of stress and its effects are linked to exposure to psychosocial hazards at work. In Europe, at least, there is a legal as well as a moral imperative to address this challenge to employees' health and safety.

The focus of this chapter has been on bringing together data on the assessment of psychosocial hazards in the work and organisational environments with that on the experience of stress and the health of employees. The case for such *triangulation* of data in the assessment of potentially stressful work and organisational environments is discussed.

Despite the advances made over the last 20 years, more research and development is required in relation to both the measurement of the experience of stress in the context of the overall stress process and to the assessment of potentially stressful work and organisational environments in terms of exposure to psychosocial hazards. A range of different methodologies have come to exist. Most are closely related to the fundamental risk assessment process. They use a variety of measures and measurement instruments. This is acceptable if the alternatives, in process or measure, are equivalent. There is an important issue of equivalence emerging.

Continuous improvement in our methods of assessment and measurement is always necessary to keep pace with changes in the nature of work, in work technologies and in organisations. The development of ever more adequate methodologies is required and they must be both evidence-based and theoretically scaffolded. What is being argued for throughout this chapter is better assessment and measurement processes, conforming to recognised good practice in relevant areas and applied within secure theoretical and methodological frameworks.

## REFERENCES

- Argote, L. and Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. *Organization Science*, 22, 1123–1137.
- Bailey, J.M. and Bagat, R.S. (1987). Meaning and measurement of stressors in the work environment. In: S.V. Kasl and C.L. Cooper (Eds.), *Stress and Health: Issues in Research Methodology*. Chichester, U.K.: Wiley & Sons.
- Burrows, G.C., Cox, T. and Simpson, G.C. (1977). The measurement of stress in a sales training situation. *Journal of Occupational Psychology*, 50, 4–51.
- Carder, M., Turner, S., McNamee, R. and Agius, R. (2009). Work-related mental ill-health and 'stress' in the UK (2002–05). *Occupational Medicine*, 59, 539–544.
- Cherry, N.M., Chen, Y. and McDonald, J.C. (2006). Reported incidence and precipitating factors of work-related stress and mental ill-health in the United Kingdom (1996–2001). *Occupational Medicine*, 56, 414–421.
- Chung, Y.-C., Park, C.-H., Kwon, H.-K., Park, Y.-M., Kim, Y.S., Doo, J.-K., Shin, D.-H., Jung, E.-S., Oh, M.-R. and Chae, S.W. (2012). Improved cognitive performance following supplementation with a mixed-grain diet in high school students: A randomized controlled trial. *Nutrition*, 28, 165–172.
- Colligan, M.J., Smith, M.J. and Hurrell Jr., J.J. (1977). Occupational incidence rates of mental health disorders. *Journal of Human Stress*, 3, 34–39.
- Cooper, C.L. (1998). *Theories of Organizational Stress*. Oxford, U.K.: Oxford University Press.
- Cooper, C.L. and Dewe, P. (2004). *Stress: A Brief History*. Oxford, U.K.: Blackwell.
- Cooper, C.L., Dewe, P. and O'Driscoll, M. (2011). *Organisational Stress: A Review and Critique of Theory, Research and Applications*. Thousand Oaks, CA: Sage Publications.
- Cox, T. (1978). *Stress*. London, U.K.: Macmillan.
- Cox, T. (1985). The nature and measurement of stress. *Ergonomics*, 25, 1129–1141.
- Cox, T. (1987). Stress, coping and problem solving. *Work & Stress*, 1, 5–14.
- Cox, T. (1993). *Stress Research and Stress Management: Putting Theory to Work*. Sudbury, Ontario, Canada: HSE Books.
- Cox, T. (1998). Work-related stress: From environmental exposure to ill-health. In: R. McCaig and M. Harrington (Eds.), *The Changing Nature of Occupational Health*. Sudbury, Ontario, Canada: HSE Books.
- Cox, T. (2011). *Psychosocial Hazards: The Risk to Employee Health of Failures in the Design and Management of Work and Work Organisations*. Canberra, Australian Capital Territory, Australia: Safe Work Australia.
- Cox, T. and Brockley, T. (1984). The experience and effects of stress in teachers. *British Educational Research Journal*, 10, 83–87.
- Cox, T. and Cox, S. (1993). *Psychosocial and Organizational Hazards: Monitoring and Control*. Occasional Series in Occupational Health, No. 5. Copenhagen, Denmark: World Health Organization.
- Cox, T. and Griffiths, A. (1995). The nature and measurement of work stress: theory and practice. In: J. Wilson and N. Corlett (Eds.), *The Evaluation of Human Work: A Practical Ergonomics*. London, U.K.: Taylor & Francis Group.
- Cox, T. and Griffiths, A. (2010). Work-related stress: A theoretical perspective. In: S. Leka and J. Houdmont (Eds.), *Occupational Health Psychology*. Chichester, U.K.: Wiley-Blackwell.

- Cox, T., Griffiths, A. and Rial-González, E. (2000). *Research on Work-Related Stress*. Luxembourg: Office for Official Publications of the European Communities.
- Cox, T., Karanika, M., Griffiths, A. and Houdmont, J. (2007). Evaluating organizational-level work stress interventions: Beyond traditional methods. *Work & Stress*, 21, 348–362.
- Cox, T. and Mackay, C.J. (1985). The measurement of self-reported stress and arousal. *British Journal of Psychology*, 76, 183–186.
- Cox, T., Oliver, A., Rial-González, E., Tomás, J.M., Griffiths, A. and Thompson, L. (2006). The development of a Spanish language version of the Worn Out Scale of the General Well-Being Questionnaire (GWBQ). *The Spanish Journal of Psychology*, 9, 94–102.
- Cox, T., Thirlaway, M. and Cox, S. (1982). Repetitive work, well-being and arousal. *Advances in the Biosciences*, 42, 115–135.
- Cox, T., Thirlaway, M. and Cox, S. (1984). Occupational well-being: Sex differences at work. *Ergonomics*, 27, 499–510.
- Cox, T., Thirlaway, M., Gotts, G. and Cox, S. (1983). The nature and assessment of general well-being. *Journal of Psychosomatic Research*, 27, 353–359.
- Crown, S. and Crisp, A.H. (1966). A short clinical diagnostic self-rating scale for psychoneurotic patients. The Middlesex Hospital Questionnaire. *British Journal of Psychiatry*, 112, 917–923.
- Cruickshank, P.J. (1982). Patient stress and the computer in the waiting room. *Social Science and Medicine*, 16, 1371–1376.
- Cruickshank, P.J. (1984). A stress and arousal mood scale for low vocabulary subjects. *British Journal of Psychology*, 75, 89–94.
- Dackert, I. (2010). The impact of team climate for innovation on well-being and stress in elderly care. *Journal of Nursing Management*, 18, 302–310.
- Denzin, N.K. (2006). *Sociological Methods: A Sourcebook*. New York: Aldine Transaction.
- Denzin, N.K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, 6, 80–88.
- Deogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H. and Convi, L. (1974). The Hopkins Symptoms Checklist (HSCL). In: P. Pichot (Ed.), *Modern Problems in Pharmacopsychiatry*, Vol. 7. Basel, Switzerland: Karger.
- Dewe, P. (1991). Primary appraisal, secondary appraisal and coping: Their role in stressful work encounters. *Journal of Occupational Psychology*, 64, 331–351.
- Dewe, P.J., O'Driscoll, M.P. and Cooper, C.L. (2012). Theories of psychological stress at work. In: R.J. Gatchel and I.Z. Schultz (Eds.), *Handbook of Occupational Health and Wellness*. Berlin: Springer.
- European Agency for Safety and Health at Work. (2010). European survey of enterprises on new and emerging risks-managing safety and health at work. Luxembourg: Office for Official Publications of the European Communities.
- European Commission. (1989). Council framework directive on the introduction of measures to encourage improvements in the safety and health of workers at work. 89/391/EEC. *Official Journal of the European Communities*, 32, 493–497.
- European Commission. (1996). *Guidance on Risk Assessment at Work*. Brussels, Belgium: European Commission.
- Evatt, D.P. and Kassel, J.D. (2010). Smoking, arousal and affect: The role of anxiety sensitivity. *Journal of Anxiety Disorders*, 24, 114–123.
- Faleel, S.-F., Tam, C.-L., Lee, T.-H., Har, W.-M. and Foo, Y.-C. (2012). Stress, perceived social support, coping capability and depression: A study of local and foreign students in the Malaysian context. *World Academy of Science, Engineering & Technology*, 61, 75–81.
- Folger, R. and Belew, J. (1985). Non-reactive measurement: A focus for research on absenteeism and occupational stress. In: L.L. Cummings and B.M. Straw (Eds.), *Organizational Behaviour*. Greenwich, CT: JAI Press.
- Goldberg, D.P. (1972). *The Detection of Psychiatric Illness by Questionnaire*. Maudsley Monograph No. 21. London, U.K.: Oxford University Press.
- Gotts, G. and Cox, T. (1990). *Stress Arousal Checklist: A Manual for Administration, Scoring and Interpretation*. Melbourne, Victoria, Australia: Swinburne Press.
- Griffiths, A. (1999). Organizational interventions: Facing the limits of the natural science paradigm. *Scandinavian Journal of Work, Environment & Health*, 25, 589–596.
- Griffiths, A.J. (1999). Organizational interventions: Facing the limits of the natural science paradigm. *Scandinavian Journal of Work, Environment & Health*, 25, 589–596.
- Griffiths, A.J., Cox, T., Karanika, M., Khan, S. and Tomás, J.M. (2006). Work design and management in the manufacturing sector: Development and validation of the Work Organisation Assessment Questionnaire. *Occupational and Environmental Medicine*, 63(10), 669–675.



- Gurin, G., Veroff, J. and Field, S. (1960). *Americans' View of Their Mental Health*. New York: Edinburgh.
- Hacker, W. (1978). *Allgemeine Arbeits- und Ingenieurpsychologie (General Work and Engineering Psychology)*. Bern, Switzerland: Huber.
- Hart, P.M., Wearing, A.J., Conn, M. and Carter, N.L. (2000). Development of the school organisational health questionnaire: A measure for assessing teacher morale and school organisational climate. *British Journal of Educational Psychology*, 70, 211–228.
- Hart, P.M., Wearing, A.J. and Headey, B. (1995). Police stress and well-being: Integrating personality, coping and daily work experiences. *Journal of Occupational and Organizational Psychology*, 68, 133–156.
- Health and Safety Executive. (2001). *Tackling Work-Related Stress: A Managers' Guide to Improving and Maintaining Employee Health and Well-Being* (No. HSG218). London, U.K.: HSE Books.
- Heaney, D., Gorman, D. and Porter, M.N. (1998). Self-recorded stress levels for general practitioners before and after forming an out-of-hours primary care centre. *British Journal of General Practice*, 48, 1077–1078.
- Ho, J. (1996). *School Organisational Health and Teacher Stress in Singapore*. Unpublished PhD thesis. University of Nottingham, Nottingham, U.K.
- Hodgson, J.T., Jones, J.R., Elliott, R.C. and Osman, J. (1993). *Self-Reported Work-Related Illness*. Sudbury, Ontario, Canada: HSE Books.
- Houdmont, J., Zhou, S. and Hassard, J. (2011). Overtime and psychological well-being among Chinese office workers. *Occupational Medicine*, 61, 270–273.
- International Labour Organisation. (1986). Psychosocial factors at work: Recognition and control. Report of the Joint ILO/WHO Committee on Occupational Health. *Occupational Safety and Health Series*, Vol. 56. Geneva, Switzerland.
- Jones, J.R., Huxtable, C.S., Hodgson, J.T. and Price, M.J. (2003). *Self-Reported Work-Related Illness in 2001/02: Results from a Household Survey*. Sudbury, Ontario, Canada: HSE Books.
- Kahn, R.L., Wolfe, D.M., Quinn, R.P., Snoek, J.D. and Rosenthal, R.A. (1964). *Organizational Stress: Studies in Role Conflict and Ambiguity*. New York: Wiley.
- Kalleberg, A.L. and Marsden, P.V. (2012). Changing work values in the United States, 1973–2006. *Social Science Research*, 42, 255–270.
- Karasek, R.A. (1979). Job demands, job decision latitude and mental strain: Implications for job redesign. *Administrative Science Quarterly*, 24, 285–308.
- Karasek, R.A. and Theorell, T. (1990). *Healthy Work: Stress, Productivity and the Reconstruction of Working Life*. New York: Basic Books.
- King, M.G., Burrows, G.D. and Stanley, G.V. (1983). Measurement of stress and arousal: Validation of the stress arousal checklist. *British Journal of Psychology*, 74, 473–479.
- Kivimäki, M., Vahtera, J., Thompson, L., Griffiths, A., Cox, T. and Pentti, J. (1997). Psychosocial factors predicting employee sickness absence during economic decline. *Journal of Applied Psychology*, 82, 858.
- Leka, S. and Cox, T. (2008). *The European Framework for Psychosocial Risk Management. PRIMA-EF*. Nottingham, U.K.: I-WHO Publications.
- Leka, S., Hassard, J., Jain, A., Makrinov, N., Cox, T., Kortum, E., Ertel, M. et al. (2008). *Towards the Development of a European Framework for Psychosocial Risk Management at the Workplace*. Nottingham, U.K.: IWHO Publications.
- Leka, S., Hassard, J. and Yanagida, A. (2012). Investigating the impact of psychosocial risks and occupational stress on psychiatric hospital nurses' mental well-being in Japan. *Journal of Psychiatric and Mental Health Nursing*, 19, 123–131.
- Leka, S. and Jain, A. (2010). *Health Impact of Psychosocial Hazards at Work: An Overview*. Geneva, Switzerland: WHO Press.
- Locke, D. and Pennington, D. (1982). Reasons and other causes: Their role in attribution processes. *Journal of Personality and Social Psychology*, 42, 212.
- Mackay, C.J. (1980). The measurement of mood and psychophysiological activity using self-report techniques. In: I. Martin and P. Venables (Eds.), *Techniques in Psychophysiology*. Chichester, U.K.: Wiley & Sons.
- Mackay, C.J., Cousins, R., Kelly, P.J., Lee, S. and McCaig, R.H. (2004). Management standards and work-related stress in the UK: Policy background and science. *Work & Stress*, 18, 91–112.
- Mackay, C.J., Cox, T., Burrows, G.C. and Lazzerini, A. (1978). An inventory for the measurement of self-reported stress and arousal. *British Journal of Social and Clinical Psychology*, 17, 283–284.
- Matthews, G., Jones, D.M. and Chamberlain, A.G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, 81, 17–42.
- McCormick, I.A., Walkey, F.H. and Taylor, A.J.W. (1985). The Stress Arousal Checklist: An independent analysis. *Educational & Psychological Measurement*, 45, 143–146.

- McHugh, M. (1997). The stress factor: Another item for the change management agenda? *Journal of Organizational Change Management*, 10, 345–362.
- Mellor, N., Smith, P., Mackay, C. and Palferman, D. (2013). The management standards for stress in large organizations. *International Journal of Workplace Health Management*, 6, 4–17.
- Mohd Mahudin, N.D., Cox, T. and Griffiths, A. (2012). Measuring rail passenger crowding: Scale development and psychometric properties. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15, 38–51.
- Nieuwenhuijsen, K., Bruinvels, D. and Frings-Dresen, M. (2010). Psychosocial work environment and stress-related disorders, a systematic review. *Occupational Medicine*, 60, 277–286.
- Quam, K.F. (2010). The mature workforce and the changing nature of work. *Research in Organizational Change and Development*, 18, 315–366.
- Ray, C. and Fitzgibbon, G. (1981). Stress, arousal and coping with surgery. *Psychological Medicine*, 11, 741–746.
- Rogers, E.H. (1960). *The Ecology of Health*. New York: Macmillan.
- Ruey-Fa, L. (1994). *Organisational Healthiness, Stress and Wellbeing*. Unpublished PhD thesis. University of Nottingham, Nottingham, U.K.
- Russell, J.A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology*, 37, 345–346.
- Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.
- Schilling, R.S.F. (1981). *Occupational Health Practice*. London, U.K.: Butterworth-Heinemann.
- Singh, M. and Woods, S.A. (2008). Predicting general well-being from emotional intelligence and three broad personality traits. *Journal of Applied Social Psychology*, 38, 635–646.
- Smith, A., Johal, S. and Wadsworth, E. (2000). *The Scale of Occupational Stress: The Bristol Stress and Health at Work Study*. Sudbury, Ontario, Canada: HSE Books.
- Stadler, T., Evans, P., Hucklebridge, F. and Clow, A. (2010). Associations between psychosocial state variables and the cortisol awakening response in a single case study. *Psychoneuroendocrinology*, 35, 209–214.
- Stansfeld, S. and Candy, B. (2006). Psychosocial work environment and mental health – a meta-analytic review. *Scandinavian Journal of Work, Environment & Health*, 32, 443–462.
- Stansfeld, S.A., Head, J. and Marmot, M.G. (2000). *Work-Related Factors and Ill-Health: The Whitehall II Study*. Sudbury, Ontario, Canada: HSE Books.
- Thyer, B.A. (2012). *Quasi-Experimental Research Designs*. Oxford, U.K.: Oxford University Press.
- Warr, P.B. (1990). Decision latitude, job demands and employee well-being. *Work & Stress*, 4, 285–294.
- Watts, C., Cox, T. and Robson, J. (1983). Morningness – eveningness and diurnal variations in self-reported mood. *Journal of Psychology*, 113, 251–256.
- Wood, L.A. (2011). The changing nature of jobs: A meta-analysis examining changes in job characteristics over time. Unpublished doctoral thesis. University of Georgia, Athens, GA.
- World Health Organization. (1946). *Constitution of the World Health Organisation*. Geneva, Switzerland: WHO.
- Yamaguchi, K. and Kanemitsu, S. (2011). Surgeon's stress from surgery and night duty. *JAMA Surgery*, 146, 271–278.

---

# 21 Psychophysiological Measurement of Physical and Cognitive Work

*Jung-Yong Kim, Young-Jin Cho and Seung-Nam Min*

## CONTENTS

|                                                            |     |
|------------------------------------------------------------|-----|
| History of Psychophysiology .....                          | 584 |
| Psychophysiology in Ergonomics .....                       | 585 |
| Electroencephalography .....                               | 585 |
| History .....                                              | 585 |
| Instrumentation .....                                      | 585 |
| Analysis and Interpretation .....                          | 586 |
| Electrocardiography .....                                  | 588 |
| History .....                                              | 588 |
| Instrumentation .....                                      | 588 |
| Analysis and Interpretation .....                          | 589 |
| ECG Variables .....                                        | 589 |
| Electromyography .....                                     | 590 |
| History .....                                              | 590 |
| Instrumentation .....                                      | 590 |
| Analysis and Interpretation .....                          | 591 |
| Nature of EMG Signal .....                                 | 591 |
| EMG Variables .....                                        | 591 |
| Integrated EMG .....                                       | 591 |
| Root Mean Square .....                                     | 593 |
| Mean Power Frequency (MPF) and Median Frequency (MF) ..... | 593 |
| Concept of EMG Normalisation .....                         | 593 |
| Frequency Analysis of Muscular Fatigue .....               | 594 |
| Stationarity of EMG Signals .....                          | 594 |
| Psychological Influence on EMG .....                       | 595 |
| Photoplethysmography .....                                 | 595 |
| History .....                                              | 595 |
| Instrumentation .....                                      | 596 |
| Analysis and Interpretation .....                          | 596 |
| Pulse-Wave Characterisation .....                          | 596 |
| Pulse-Wave Analysis Technique .....                        | 597 |
| Skin Conductance Response .....                            | 598 |
| History .....                                              | 598 |
| Instrumentation .....                                      | 598 |
| Analysis and Interpretation .....                          | 599 |

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| Skin Temperature .....                                             | 599 |
| Theoretical Background .....                                       | 599 |
| Electrooculography .....                                           | 599 |
| Theoretical Background .....                                       | 599 |
| Recent Studies and Issues Using Psychophysiological Measures ..... | 600 |
| Brain-Computer Interface .....                                     | 600 |
| Automobile Safety .....                                            | 601 |
| Motor Control.....                                                 | 601 |
| Conclusions.....                                                   | 602 |
| References.....                                                    | 602 |

## HISTORY OF PSYCHOPHYSIOLOGY

Psychophysiology is an area often referenced by philosophers, physicians, physiologists and psychologists. Since the 1950s, researchers working in the area of physiological psychology have been called 'psychophysiologicalists'. They were interested in psychophysiological phenomena to understand human nature and behaviour. Recently, ergonomists also have shown interest in using psychophysiology to understand and quantify the interaction of the human mind with products and the environment.

The ancient Greeks wondered where human intellect, emotion and instinct reside in the body. The philosopher Plato believed that a human is composed of three parts: the head, controlling rational function; the spinal marrow, governing passion and the heart; and the spinal cord below the diaphragm, which is related to instincts or 'base' appetites. In India, our body is perceived as a multilayered entity with 'chakras' that define psychological and physiological states depending on an individual's ability to activate potential mental energy. Practitioners of oriental medicine use a holistic view of the body, such as the theory of Yin and Yang, along with accumulated clinical experience to examine and treat patients without examinations with instruments. Likewise, intuitive analyses have been applied to studies of human psychology or physiology instead of experimental investigation.

Such an instrumental approach to the collection of psychophysiological information was noted by Erasistratus (304–250 BC), a Greek anatomist and royal physician. While treating the serious illness of a general's son who fell in love with his stepmother, he showed that physical ailments might be closely related to mental problems. Using a technique similar to a lie detector, Erasistratus observed 'stammering speech, sudden sweats, irregular palpitations of the heart' in the man when his stepmother came into the room and was facing him. This experimental story was recorded by Plutarch (AD 46–120), a Greek historian. Mesulam and Perry (1972) described Erasistratus as an early psychophysiologicalist who founded the theory of stimulus–response specificity. Galen (AD 129–200), the famous Roman physician who is remembered as the father of modern physiology, examined a patient suffering from lovesickness by observing her irregular pulse when she heard the name of her lover. Ibn Sina (Avicenna) (AD 980–1037), a Persian philosopher, scientist and author of *The Book of Healing*, also used the pulse rate to evaluate whether patients were in love.

In the eighteenth century, Luigi Galvani (1737–1798), an Italian physician and physicist, performed an important experiment to develop psychophysiology further (Hoff, 1966). Galvani experimentally demonstrated that animals could generate electricity in their organs. Coupled with past findings in which frog leg muscles were shown to contract when electricity was applied, Galvani developed a medical hypothesis suggesting that disease can be detected by observing the pattern of electrical current in the body. The electrical nature of muscle contraction became the fundamental basis for various measuring devices developed in recent psychophysiology studies.

## PSYCHOPHYSIOLOGY IN ERGONOMICS

Ergonomics/human factors (E/HF) is the scientific discipline concerned with the interactions among humans and the other elements of a system and involves the application of theory, principles, data and methods to optimise overall system performance as well as human well-being (see Chapter 1). As part of their mission to accomplish this goal, ergonomists have searched for methodologies to find the relationship between quantifiable measures and physical and mental workloads. Thus, psychophysiology has become one of the most important methodologies for understanding mental behaviour and human interactions with system components. This chapter is designed to introduce existing techniques in psychophysiology to a wide variety of readers who want to utilise the methods to assess human behaviour in a quantitative manner. The measures presented are those most frequently used in human interaction studies, and the utility of the measures is only partially introduced in this chapter. Most of the information in this chapter was referenced from various publications and reformatted for easy understanding. The measurement techniques covered in this chapter are electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG), photoplethysmography (PPG), skin temperature (SKT), skin conductance response (SCR) and electrooculography (EOG). In addition, recent research issues including various applications of psychophysiology are discussed at the end of the chapter.

## ELECTROENCEPHALOGRAPHY

### HISTORY

EEG (also discussed in Chapter 22 within the specific context of neuroergonomics) is a knowledge domain encompassing the recording and interpretation of electroencephalograms (EEGs), which capture electrical activity from the scalp area. EEG measured directly on the cortex is known as an electrocorticogram (ECoG) (Tomarken et al., 2007). EEG measures voltage fluctuations originating from ionic current flows in neurons and especially extracellular field potentials generated by the synchronous activities of a great number of neurons. The word ‘electroencephalography’ is derived from the Greek words *enkephalo* (brain) and *graphein* (to write).

In one of the most remarkable developments in the history of clinical neurology, Hans Berger invented the EEG. Berger successfully recorded a human EEG for the first time in 1924. He published his first paper in 1929 describing the methodology for ‘recording the electrical activity of the human brain from the surface of the head’ (Millet, 2002).

EEG has been used to measure human emotion or sensibility (Elul, 1972; Fox, 1991; Klados, 2014). Davidson and Fox (1982) studied the reaction of 10-month-old infants watching video tapes. As sad and happy scenes were shown, laterality in the EEG signals was noted. That is, happy and pleasurable emotion activated the left frontal lobe, whereas sad and unpleasant emotion activated the right and left parietal lobe.

EEG recorded without an external stimulus is called spontaneous EEG; EEG generated by external or internal stimulus measures an event-related potential (ERP). The range of EEG amplitude recorded from the scalp area among normal participants is 10–100  $\mu$ V (Torse et al., 2012). In the cortex, the amplitude range is 500–1500  $\mu$ V (Blinowska and Durka, 2006).

### INSTRUMENTATION

The voltage measured by EEG equipment is derived from the difference in the electrical charge measured over the cortical area of interest and that at the reference electrode. The diameter of the electrodes is around 6–8 mm. The analogue-to-digital sampling rate is typically 256–512 Hz in clinical scalp EEG. For research purposes, the sampling rate can be between 250 and 2,000 Hz, but

newly developed data collection systems can record data above 20,000 Hz. Traditionally, the scalp area is prepared by rubbing it with cleansing alcohol on a cotton ball until the skin shows a tinge of pink. Then, an electrode paste is carefully placed on the skin of the scalp to produce a certain level of conductivity. Despite these procedures to improve skin conductivity, it may be difficult to achieve a good signal-to-noise ratio to collect detailed information regarding neural activity. To overcome this, the electrodes need to be attached firmly using an elastic headband with Velcro or other commercially available devices. Many researchers use 'electrode caps' with prepositioned electrodes embedded at locations based on the international 10–20 system for EEG measurement (Herwig et al., 2003) (described in later section).

For EEG electrodes, gel-type electrodes have been widely used compared to dry electrodes due to the sensitivity and price. However, a dry electrode is now used in many EEG devices because of convenience of not using conductive gel. Taheri et al. (1990) introduced single and multichannel dry active electrode arrays. In human participants, he tested the electrode in four modalities of EEG activity, including spontaneous EEG, sensory ERPs, brain stem potentials and cognitive ERPs. The performance of dry electrodes was favourable compared to standard wet electrodes in terms of skin preparation, absence of gel requirements and higher signal-to-noise ratio (Taheri et al., 1990). For these reasons, dry electrodes have been used widely in assistive devices for handicapped individuals and in various gaming applications. However, the reliability of the signals has not been validated fully in many cases. Gargiulo et al. (2014) discuss that dry electrodes allow more user-friendly and pervasive patient monitoring, but proof is required that new devices can perform biopotential recording with a quality at least comparable to existing medical devices.

EEG can be conducted using a variety of EEG helmets with up to 256 built-in electrodes, and it is also still possible to use only two electrodes. Of course, one can record from as many electrodes as possible depending upon the research questions that are being asked. However, if one is only interested in EEG responses associated with particular behaviour, one may choose the regions of the scalp that are known to be or predicted to be associated with the target behaviour.

The international 10–20 system is the standard methodology for electrode placement. In this system, letters stand for areas of the brain. The odd numbers indicate the left side and the even numbers indicate the right side (Figure 21.1). That is, C3 means that the electrode is placed on the central region of the brain on the left side, whereas C means that an electrode is placed at the top of the scalp in the central area. These lead locations are conventional; however, non-standard electrode locations may be used to investigate a particular research question.

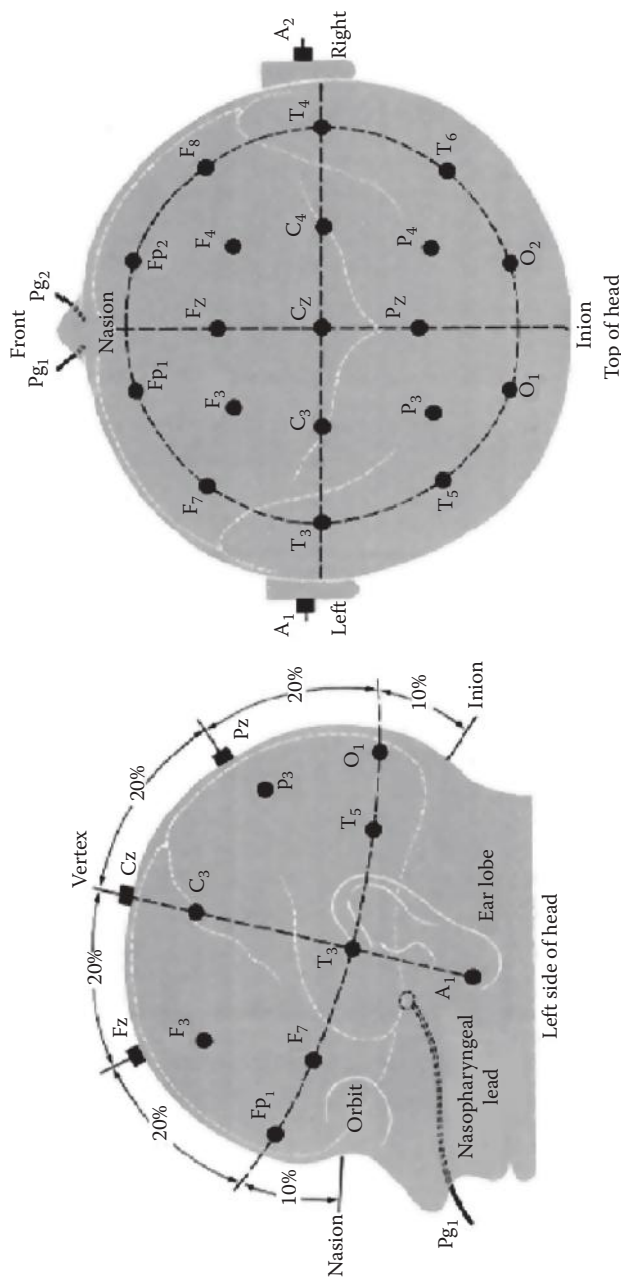
## ANALYSIS AND INTERPRETATION

Although individual variability is often observed in psychophysiological responses, the characteristics of EEG recordings, including frequencies and amplitudes, are relatively consistent within the same individual. Tomarken et al. (1992) reported a high correlation between alpha waves recorded 3 weeks apart in the same subject. Stassen et al. (1988) also found similarity between the EEG patterns of identical twins that had been raised apart. Further, they reported a similarity in the EEG patterns of fraternal twins compared to unrelated persons.

The reliability of EEG patterns produced in response to various environmental stimuli is relatively high. In particular, the frequency attribute of EEG recordings has been shown to provide referential information for understanding a participant's mental state, including comfort, tension, relaxation and drowsiness. These frequency patterns are denoted by the Greek letters  $\delta$  (delta),  $\theta$  (theta),  $\alpha$  (alpha),  $\beta$  (beta) and  $\gamma$  (gamma). The upper and lower limits of these frequency ranges vary somewhat by study (Table 21.1). Some researchers define the bands using decimal values rather than rounding values, and many scholars choose their own target frequency bands.

Delta waves appear during deep sleep in adults. This is also the predominant frequency of babies. Theta waves occur in young adults during drowsy, aroused or physically idle states. Alpha waves are easily detected when one is resting quietly in a relaxed position, with eyes closed. Beta waves





**FIGURE 21.1** The international 10–20 electrode system (O, occipital; P, parietal; C, central; F, frontal; and T, temporal). (From Jasper, H.H., *Electroencephalogr. Clin. Neurophysiol.*, 10, 371, 1958.)

**TABLE 21.1**  
**Range of Brain Waves**

| Area               | Frequency (Hz)               | Frequency (Hz)                | Frequency (Hz)      |
|--------------------|------------------------------|-------------------------------|---------------------|
| Delta ( $\delta$ ) | 0.5–3.5                      | 0.5–4                         | 0.5–4               |
| Theta ( $\theta$ ) | 3.5–7.5                      | 4–7                           | 4–8                 |
| Alpha ( $\alpha$ ) | 7.5–12.5                     | 8–13                          | 8–12                |
| Beta ( $\beta$ )   | 12.5–35                      | 13–30                         | 18–30               |
| Gamma ( $\gamma$ ) | 35–100                       | Above 30                      | 30–70 or higher     |
| Researchers        | Schneider and Strüder (2012) | Constant and Sabourdin (2012) | Stern et al. (2001) |

are shown when a person is mentally or physically excited or alert. Gamma waves can be observed when the brain is used to integrate a variety of stimuli such as sound and sight into a coherent whole (Kisley and Cornwell, 2006; Kanayama et al., 2007).

## ELECTROCARDIOGRAPHY

### HISTORY

ECG (also known as EKG) is a non-invasive measure to record the electrical activity generated by the heart over a period of time. ECG records and displays the cyclic waves of the heartbeat.

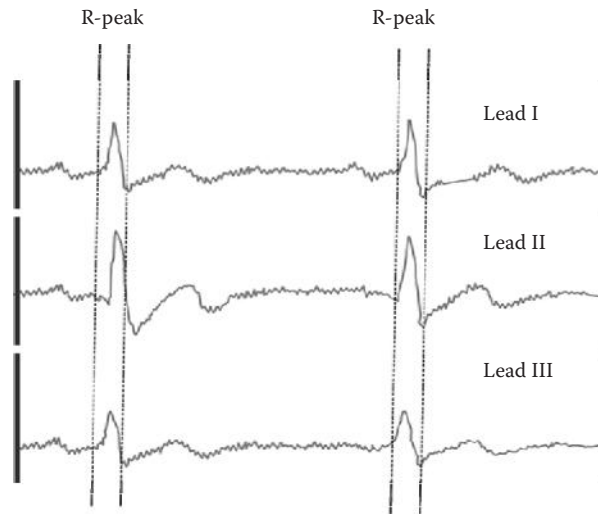
In 1872 at St. Bartholomew's Hospital, Alexander Muirhead, who was working on his doctor of science degree, connected wires to a patient's wrist to record the heartbeat (Ronald, 2004). The British physiologist John Burdon Sanderson visualised this electrical activity using a Lippmann capillary electrometer. However, the first systematic and electrical measurement of the heartbeat was done by Augustus Waller at St. Mary's Hospital in Paddington, London. He photographically projected the trace of the heartbeat and recorded it in real time (Waller, 1887). In 1903, Willem Einthoven, a Dutch doctor and physiologist, invented the first practical ECG, for which he received the Nobel Prize in Medicine in 1924. Einthoven's string galvanometer was continuously used until the advent of vacuum tube electrocardiographic instrumentation in the 1930s and 1940s.

### INSTRUMENTATION

Initially, a container of saline solution, into which subjects immersed their arms and legs, was used as an ECG electrode. Nowadays, silver–silver chloride electrodes are commonly used for ECG recording. A standardised electrode placement system on the limbs and chest is used for medical diagnosis. The basic requirement for the measurement is a good-quality heartbeat signal at two electrodes that are located far apart. Usually, 10 electrodes are used in a '12-lead' ECG. The term 'lead' refers to the electrical cable attaching the electrodes to the ECG recorder.

Figure 21.2 shows ECG signals recorded simultaneously from four leads on one subject. Leads I and III are less important for psychophysicologists because they show smaller 'R peak amplitude', which is the main feature used for arrhythmia detection (see the first and second channels). Lead II and modified lead II (right collarbone and lower left rib) are more meaningful for psychophysicologists because they provide information about a pronounced R spike.

The standard procedure for experiments is as follows. The participant sits comfortably on a chair. The experimenter gives the subject a brief overview of the purpose of the experiment. It is important to explain how the heart rate (HR) is measured and to inform the participant that there is no risk of electric shock or any penetration of electricity through the electrodes. To lower the electrical resistance, the target skin area is cleaned and rubbed with alcohol to remove oils or dead skin.



**FIGURE 21.2** ECGs derived from standard limb leads I, II and III.

Standard silver–silver chloride electrodes and conductive gel are used. An electrode cup with the electrode is attached to the skin with an adhesive collar or tape. Disposable electrodes, gelled by the manufacturer for immediate application, are now commonly used in laboratories.

### ANALYSIS AND INTERPRETATION

Because the ECG records a relatively large amplitude wave, it does not require as much amplification as EMG or EEG. Even with low amplification, we can pick up appropriate electrical activity. Stern et al. (1980) described that a filter can be set at 20–30 Hz because the HR has a frequency of about 1 Hz. In this case, 60 Hz noise and most muscle artefacts are not detected in ECG.

In the early days of ECG study, the distance between R waves was recorded on paper and measured with a ruler to calculate the HR. Researchers now use either a cardiometer or an electronic device to measure the timing between beats. The most common measures are HR or heart period (HP).

### ECG Variables

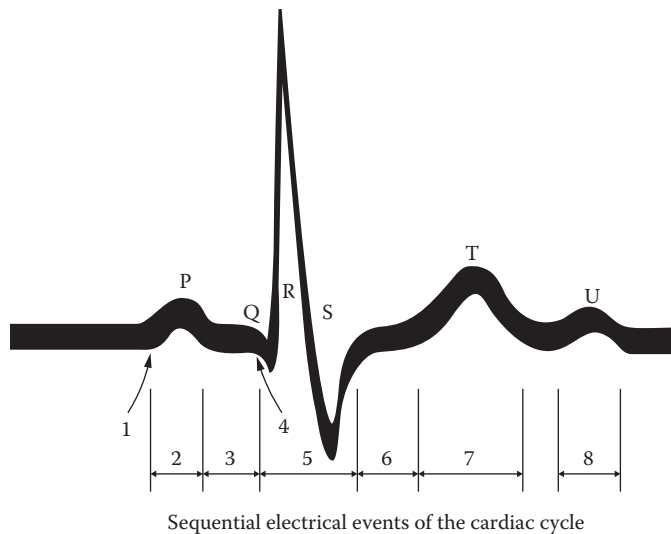
#### *Beats per Minute or Heart Rate*

The HR is based on the number of beats per unit of time. The beats per minute (BPM) is commonly used. In this case, the R wave is an indicator of the heartbeat (Figure 21.3). Thus, a continuous and sufficient sample is necessary to compute BPM. The heartbeat sample can be collected for 10–20 s to estimate BPM.

#### *Heart-Rate Variability*

The heart rate variability (HRV) is recognised as an important psychophysiological variable (Berntson, 1997). It is measured in terms of the variation of beat-to-beat interval. In practice, the coefficient of variance is used as a measurable indicator for HRV. Other variables such as cycle length variability or RR (interval between successive Rs) variability can be used. It is known that HRV is closely associated with emotional arousal. For example, it has been found that high-frequency activity decreases under acute time pressure and emotional strain (Nickel and Nacheriner, 2003) and an elevated state of anxiety (Jönsson, 2007).

HR and HRV are not independent of each other due to their dependency on the same autonomic nervous systems (ANS). HR forms the best index of the ANS balance and is the easiest to measure,



**FIGURE 21.3** Typical ECG and description of sequential electrical events of the cardiac cycle: (1) impulse from the sinus node, (2) depolarisation of the atria, (3) depolarisation of the A-V node, (4) repolarisation of the atria, (5) depolarisation of the ventricles, (6) activated state of the ventricles, (7) repolarisation of the ventricles and (8) after-potentials following repolarisation of the ventricles. (From Phillips, R. and Feeney, M., *The Cardiac Rhythms: A Systematic Approach to Interpretation*, W. B. Saunders Co., Philadelphia, PA, 1973.)

and all other indices including HRV are influenced by HR. Therefore, the HR is the important variable to be considered when HRV is evaluated (Coulmel et al., 1995).

## ELECTROMYOGRAPHY

### HISTORY

EMG records electric activity generated by muscles during contraction and is used for the diagnosis of neuromuscular disease and the estimation of muscle force or local muscle fatigue. EMG information is often used for kinetic and kinematic analyses of human motion.

In 1666, the first documented experiments related to electric muscle activity were done by Francesco Redi (1626–1697), an Italian physician, naturalist and poet. Redi found that the electric eel had a special muscle capable of generating electricity. Luigi Galvani wrote a book in 1771 entitled *De Viribus Electricitatis in Motu Musculari Commentarius* in which he noted that muscle contraction can be initiated by electricity. In 1849, the first recording of surface EMG signals was performed by Raymond, and in 1944, Inmann et al. first reported the activity of different shoulder muscles by analysing surface EMG signals. Recently, EMG signal detection, amplification and processing systems have been improved and computerised, and increasing amounts of quantitative and qualitative information can be extracted from motor-unit action potentials (MUAPs) through EMG. The ease with which EMG devices can be operated has greatly increased the number of users. EMG is currently used mostly by researchers and clinical practitioners, in fields including neurophysiology, biomechanics, rehabilitation, sports science and space medicine.

### INSTRUMENTATION

When EMG is used in the clinic to examine or study muscular disorders, very thin needle electrodes are inserted into muscle tissue. In this case, the electrical signal of muscle activity is collected from limited muscle regions or even from a single motor unit. EMG can also be

recorded using surface electrodes. In this case, some portion of the action potentials generated by muscle motor units is transmitted to surface electrodes through the skin surface. Most EMG studies related to human performance are associated with large muscles. Therefore, surface electrodes represent a safe and easy choice that provides the necessary information about muscle performance.

The methods of electrode application are the same as with other physiological measures. That is, the skin needs to be rubbed with alcohol to remove dead cells, dirt and oils. Usually, a conductive gel or paste is used to create a contact between the skin and the surface electrode. Pre-gelled surface electrodes have been used widely because they are easily attached and reduce the duration of experiments. Recently, dry electrodes in wireless EMG and EEG systems have been commercially developed and used for convenient electrode attachment (Wheeler and Jorgensen, 2003; Forvi et al., 2012).

It is recommended that the resistance between the electrodes be less than 10,000  $\Omega$ . In cases where the resistance is less than 1,000  $\Omega$ , unnecessary conduction between electrodes can be detected if the electrode gel or paste is overused or if the electrodes are placed too closely together. In bipolar placement, the electrode array usually consists of two active and one ground electrode. The active electrodes are placed in a bipolar pattern along the long axis of the target muscle. The amplitude of EMG is the algebraic sum of all action potentials of the contracting motor units located between the electrodes. The specifications for electrode placement on different muscles have been summarised by Davis (1959) and Basmajian and Blumenstein (1980). Eleanor (2011) showed the anatomical positions of selected electrode sites for surface EMG (Figure 21.4).

## ANALYSIS AND INTERPRETATION

### Nature of EMG Signal

An unfiltered (with the exception of high- and low-pass filters) and unprocessed signal is called a raw EMG signal. Figure 21.5 shows a raw surface EMG signal (sEMG) recorded during three static contractions of the erector spinae muscle.

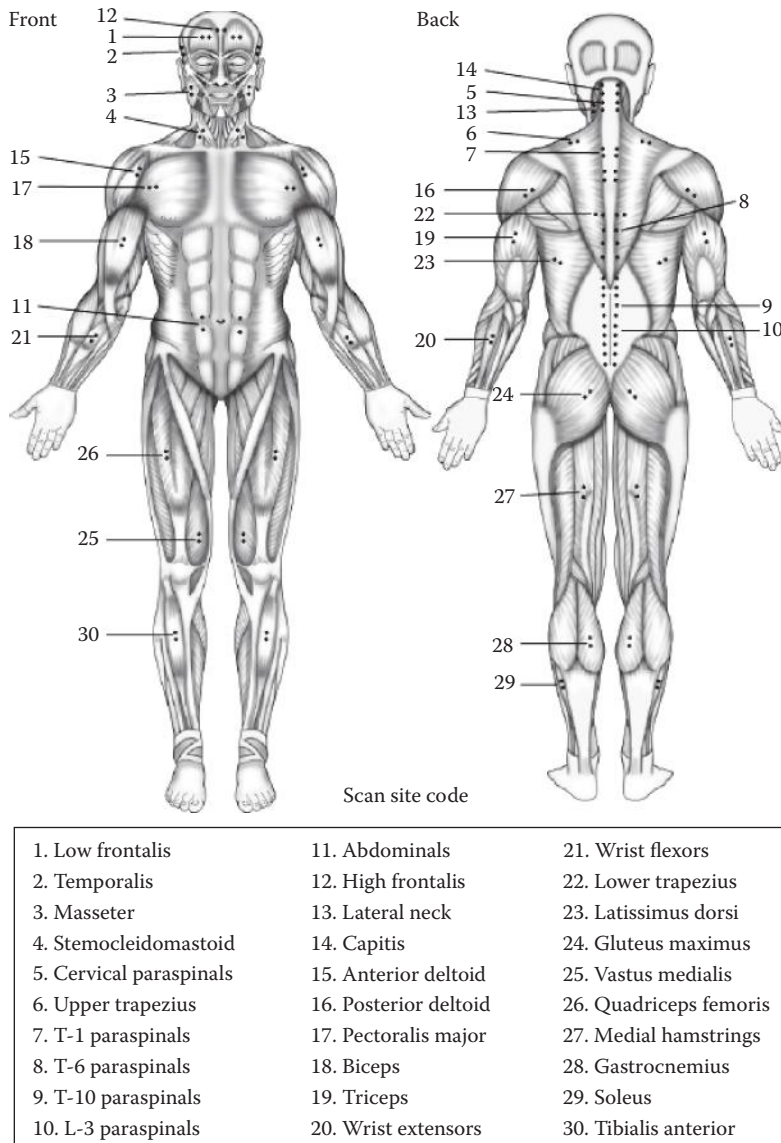
The baseline raw EMG can be observed when the muscle is relaxed. However, raw EMG baseline noise is often observed due to the poor quality of the EMG amplifier and/or wire and the electric field of alternating current, for example. Assuming state-of-the-art EMG hardware performance and proper skin impedance, the averaged baseline noise is usually less than 3–5  $\mu\text{V}$  (Gheab and Saleem, 2008). The examination of the baseline EMG signal is very important for reliable EMG measurement. The experimenter should be able to differentiate the pattern of ‘increased’ baseline activity and abrupt muscle contraction.

## EMG VARIABLES

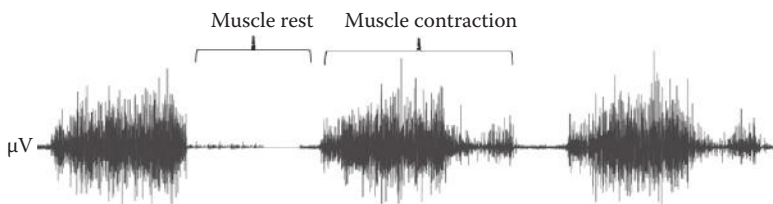
The most commonly used parameters in the time domain are *integrated EMG* (IEMG) and *root mean square* (RMS). In the frequency domain, the mean power frequency (MPF) and median frequency (MF) are usually used to analyse the frequency contents of the EMG signal. To compute these parameters, a fast Fourier transform (FFT) algorithm is usually used.

### Integrated EMG

The basic step of EMG signal analysis is rectification. All negative amplitudes of the signals are converted to positive amplitudes. After the rectification of the signals, IEMG can be calculated. IEMG is defined as the area under the curve of the rectified EMG signal, that is, the mathematical integral of the absolute value of the raw EMG signal. IEMG splits the signal into fixed-width time slices and resets the integral at the start of each time slice. To derive IEMG, the width of this time slice must



**FIGURE 21.4** Scan sites for the entire body. (From Criswell, E., *Cram's Introduction to Surface Electromyography*, Jones & Bartlett Publishers, Sudbury, MA, 2011.)



**FIGURE 21.5** Raw EMG recording of three contractions in the erector spinae muscle.



be specified ([www.biopac.com](http://www.biopac.com)). However, researchers must consider the constant increase of the mathematical integral value due to noise when the absolute value of the signal is taken:

$$I\left\{\left|EMG_{(t)}\right|\right\} = \int_0^t |EMG_{(t)}| dt.$$

Here  $t$  is the fixed time interval (U.S. Department of Health and Human Services, 1992).

### Root Mean Square

The RMS of EMG is based on the square root of data. This method does not require rectification of the signal:

$$RMS\{EMG(t)\} = \frac{1}{T} \left( \int_t^{t+T} EMG^2(t) dt \right)^{1/2}.$$

This algorithm uses the time window, and the window size depends on the contraction time associated with the movement speed. The shorter the time window used, the smoother the EMG pattern that can be extracted.

### Mean Power Frequency (MPF) and Median Frequency (MF)

MPF is the average of all frequencies. MF is the frequency having 50% of the frequency distribution on each side. The MF appears to be less sensitive to noise than the MPF:

$$f_{\text{mean}} = \frac{\int_0^f f S_m(f) df}{\int_0^f S_m(f) df},$$

$$\int_0^{f_{\text{median}}} S_m(f) df = \int_{f_{\text{median}}}^{\infty} S_m(f) df$$

where

$f_{\text{mean}}$  is the mean power frequency

$f_{\text{median}}$  is the median frequency

$S_m(f)$  is the power density function (Basmajian and Luca, 1985)

In medical practice or academic research, IEMG and/or RMS of EMG are used to measure the on–off state of neuromuscular activity and estimation of muscle force generation, whereas the MPF and MF are used to evaluate the level of muscular fatigue. The performance of various fatigue measures was compared with each other in a controlled study (Kim et al., 2005). MF was found to be more sensitive and less reliable than MPF to show the rate of fatigue over 20 s period at 75% MVC during isometric exertion of the erector spinae muscle. The difference between two parameters was not significant in the study.

### CONCEPT OF EMG NORMALISATION

Normalisation is a procedure in which the absolute EMG values are expressed as a percentage of a reference EMG value obtained during a maximal or submaximal contraction test. The most commonly used reference value is the maximal myoelectric activity, which is elicited by the maximum voluntary isometric contraction (MVC). EMG data expressed relative to the maximum (% MVC)

have the advantage of having physiological relevance; however, we have to use the submaximal reference when MVCs are limited by aging, pain or other symptoms (Allison et al., 1998; Marras and Davis, 2001; Dankaerts et al., 2004). However, for rehabilitation purposes, acceptable maximum effort (AME) can be used for biofeedback-based treatment. The equation to calculate the normalised EMG activity was suggested by Seroussi and Pope (1987) and Marras et al. (1987):

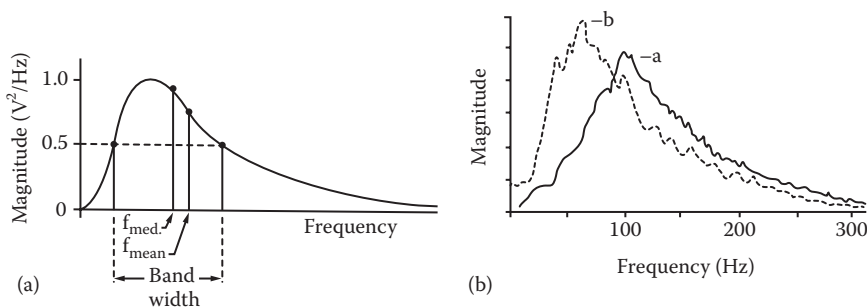
$$\text{Normalised EMG}(\%) = \frac{\text{EMG}_{\text{Task}} - \text{EMG}_{\text{Rest}}}{\text{EMG}_{\text{Max}} - \text{EMG}_{\text{Rest}}} \times 100.$$

## FREQUENCY ANALYSIS OF MUSCULAR FATIGUE

A frequency analysis of EMG signals is necessary to assess local muscle fatigue. The frequency content of EMG signals provides information about motor-control strategies, muscle fibre-type compositions and the physiological mechanisms of fatigue (Lindström and Magnusson, 1970; Broman et al., 1985; Solomonow et al., 1990; Gerdle et al., 2000). That is, the different characteristics of the slow-twitch fibre, which is known to be fatigue-resistant, and the fast-twitch fibre, known to be force-generating, induce the median or mean frequency shift towards a lower frequency band when the local muscle experiences fatigue (Figure 21.6). This frequency analysis technique is a sensitive measure for assessing the impending fatigue of local muscles given a certain workload. Therefore, a spectral analysis including the frequency contents of EMG signals is a useful tool for assessing the level of fatigue in ergonomic studies. Various commercialised devices with analysis software are available for researchers and practitioners who are not familiar with the FFT technique used in frequency analysis.

## STATIONARITY OF EMG SIGNALS

FFT and the discrete Fourier transform (DFT) used in spectral analysis have become the most widely used mathematical algorithms in EMG analyses (Petrofsky, 1979; Horita and Ishiko, 1987; Perry et al., 2001; Beck et al., 2004; Coburn et al., 2005). To apply these algorithms to EMG signals, it is necessary to satisfy the basic assumption of the Fourier transform, which is the stationarity of the signal (Oppenheim and Schaffer, 1989; Blanco et al., 1995; Bilodeau et al., 1997). Stationarity is the quality in which statistical parameters such as mean and standard deviation do not change with time. A strong stationary signal indicates that all statistical properties are



**FIGURE 21.6** Frequency analysis of EMG signals: (a) the median and mean frequency in frequency spectrum of EMG signals and (b) the frequency shifts resulting from local muscle fatigue during a sustained, isometric contraction of the first dorsal interosseous muscle. (a) Frequency distribution of EMG signal before fatiguing and (b) frequency distribution of EMG signal after fatiguing. (From the U.S. Department of Health and Human Service, *Selected Topics in Surface Electromyography for Use in the Occupational Setting: Expert Perspective*, Public Health Service.)

time-invariant, whereas a weakly stationary signal shows that only some statistical properties, such as mean and the autocorrelation function, are time-invariant (Cho and Kim, 2012).

However, most authors have not been concerned with the stationarity of an EMG signal even during a spectral analysis (Beck et al., 2006). The importance of signal stationarity has been discussed and tested by many authors (Balestra et al., 1988; Lo Conte and Marlette, 1995; Merlette and Lo Conte, 1997; Farina and Marlette, 2000; Beck et al., 2006; Cho and Kim, 2012). Coorevits et al. (2008) stated that most fatigue studies of the back muscle under isometric conditions presumed signal stationarity without actually testing it, which may lead to a low-quality analysis of the signal.

Some researchers have suggested appropriate window sizes for the various experiments to assure the signal stationarity of EMG. EMG signals have been shown to be stationary with 500 and 1000 ms window sizes during 50% MVC (Inbar and Noujaim, 1984); 60–1000 ms window sizes during 25% MVC (Popivanov and Todorov, 1986); a 640 ms window size during 1 kg weight loading (Paiss and Inbar, 1987); 500–1500 ms at 50%–80% MVC (Merletti et al., 1992); and a 750 ms window size at 0%–50% MVC during isometric trunk extension at various load, flexion and twisting conditions (Cho and Kim, 2012). Moreover, an EMG signal can be non-stationary during dynamic contraction because of the changes in muscle length, muscle force, velocity and the location of electrodes (Roy et al., 1998; Knaflitz and Bonato, 1999).

Recently, time–frequency methods such as short-time Fourier transform (STFT) and wavelet transform that can analyse the non-stationary signals without assuming signal stationarity have been used for spectral analyses (Karlsson et al., 2000; Akataki et al., 2001; MacIsaac et al., 2001; Beck et al., 2005). The STFT breaks the signal into overlapping observation windows (segments) that may demonstrate the local stationarity of the signal. In particular, the STFT performs well when the properties of the non-stationary signal change slowly. The wavelet transform does not assume signal stationarity and uses different length windows (wavelets) to achieve a multiresolution analysis. These methods can estimate the power spectrum of non-stationary EMG signals. The wavelet transform also performs well for stationary EMG signals (Hostens et al., 2004). Many time–frequency approaches, however, require a significant amount of computation time and a compromise between time and frequency resolution (Bonato et al., 1996; Karlsson, 2000; Beck et al., 2005).

## PSYCHOLOGICAL INFLUENCE ON EMG

Marras et al. (2000) studied the relationship between psychosocial stress and muscle activities. They evaluated the influence of psychosocial stress on the functioning of the biomechanical system and subsequent spine loading and reported that psychosocial stress increases spine compression and lateral shear. They suggested a potential pathway between psychosocial stress and spine loading. Schleifer et al. (2008) examined the influence of mental stress on muscle activity during a computer-based task to find psychophysiological mechanisms by which psychosocial stress factors contribute to musculoskeletal problems. They observed that the effect of momentary rest after musculoskeletal work was diminished by mental stress. Nimbarte et al. (2012) explored the influence of individual personality on the biomechanical response of neck and shoulder muscles to physical and psychosocial demands. They indicated that mental stress during physically demanding tasks seemed to be mediated by the individual's personality.

## PHOTOPLETHYSMOGRAPHY

### HISTORY

Because the volume of circulating blood greatly affects the amount of oxygen delivered to important body parts, psychophysiologicalists have been interested in measurements of blood circulation, obtained by observing HR, blood pressure, blood volume and blood flow.

The most commonly used tool to determine blood volume or circulation is PPG. The term 'plethysmography' derives from the Greek word meaning 'enlargement' or 'fullness'. Technically, a plethysmograph is a device for measuring the change in volume of a given body part such as the arm, leg, finger or even the entire body. Three types of measurements are commonly used: photoelectric plethysmography, impedance plethysmography and strain-gauge plethysmography. Pulse oximetry is most often used in ergonomic investigations; this non-invasive method measures the saturation level of haemoglobin in tissue. A sensor is placed on a thin part of body, usually a fingertip or earlobe. PPG or a pulse oximeter measures the saturation of peripheral oxygen ( $\text{SpO}_2$ ) level.

In 1935, Karl Matthes (1905–1962), a German physician, developed the first two-wavelength ear  $\text{O}_2$  saturation metre with red and green filters, which were later switched to red and infrared filters. His was the first device to measure  $\text{O}_2$  saturation (Matthes, 1935).

The original oximeter was made by Glenn Allan Millikan in the 1940s; in 1949, Wood added a pressure capsule to obtain the absolute  $\text{O}_2$  saturation value. In 1964, Shaw assembled the first absolute ear oximeter. The device was commercialised by Hewlett–Packard, but its use was limited to pulmonary functions and sleep laboratories due to its high cost and large size. In 1972, Takuo Aoyagi and Michio Kishi, bioengineers at Nihon Kohden, developed a pulse oximeter that evaluated the ratio of red to infrared light absorption. It was commercialised by Biox in 1981 and Nellcor in 1983 for wide distribution.

## INSTRUMENTATION

Photoelectric PPG uses a photocell placed over an area of tissue perfused with blood. One method measures the transmission of a light source passing through the tissue segment; the second measures the reflection of the light source off the tissue. The infrared light (wavelengths ranging from 7000 to 9000 Å) that passes through or is reflected is captured by a photoelectric transducer and converted into an electrical signal.

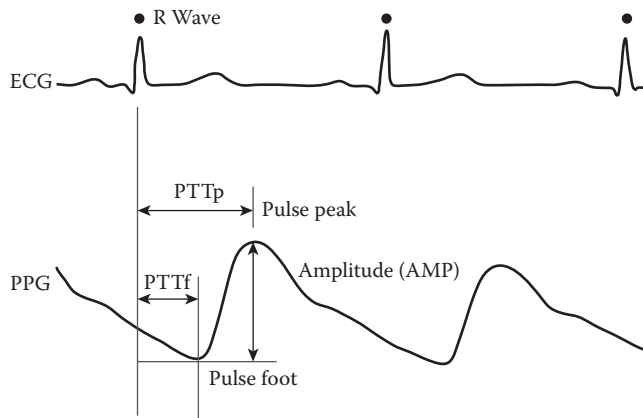
The transmission mode generally is more restrictive than the reflection mode because of factors such as ambient light interference. Thus, the PPG probe should be fastened securely in place to minimise probe–tissue movement artefacts. The artefact can be reduced by shading the measured area and obtaining measurements under subdued lighting and by electronic filtering (e.g. light-modulation filtering) (Webster, 1997). Ambient light interference in PPG-based systems has also been discussed by Hanowell et al. (1987).

Currently, most PPG devices use either a light-emitting diode (LED) or a phototransistor that is safe enough not to affect skin and the blood vessels being monitored. The pulse oximeter used on fingertips obtains the absorption change by subtracting the minimum absorption from the peak absorption (Webster, 1997). A monitor can discern only the absorption caused by arterial blood, ignoring other tissues or nail polish. Filtering circuitry is also needed to remove unwanted higher-frequency noise, such as electric frequency interference. Recently, remote PPG imaging technology has been developed. In 2005, Wieringa et al. described a contactless multiple wavelength PPG imaging system with a primary application in the remote imaging of the  $\text{SpO}_2$  distribution within tissue. PPG is lightweight, easily installed and easy to use and is capable of robust data processing, giving it the potential to become a ubiquitous medical device for remote/home health monitoring of patients. To improve the system, sensor placement should be optimised and techniques to reduce motion artefacts and minimise battery power consumption should be considered.

## ANALYSIS AND INTERPRETATION

### Pulse-Wave Characterisation

Two important characteristics of the PPG AC pulse waveform were described by Hertzman and Spealman (1937). The pulse was defined as having two phases. The first phase, called the acrotic phase, occurs at the rising edge of the pulse and is primarily concerned with systole;

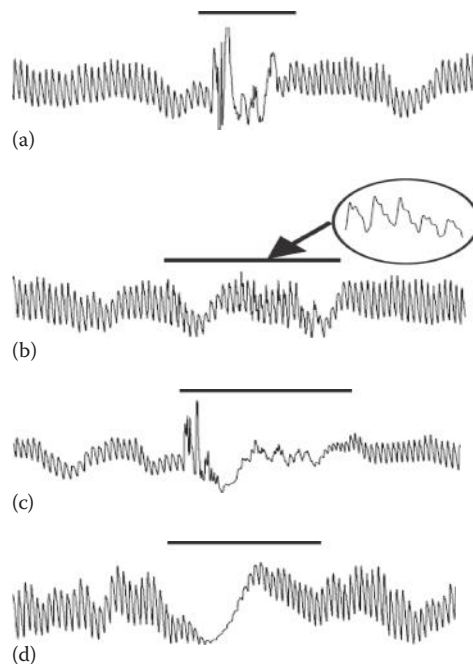


**FIGURE 21.7** Characterisation of PPG pulse timing, amplitude and shape features. Key pulse landmarks are identified and used to compute the beat-to-beat pulse transit time to the foot of the pulse (PTTf), pulse transit time to the peak of the pulse (PTTp) and foot-to-peak amplitude (AMP). (From Allen, J. and Murray, A., *Physiol. Measure.*, 21(3), 369, 2000.)

the second phase, the catacrotic phase, occurs at the falling edge of the pulse and is related to diastole. The pressure pulse wave changes its shape temporally as it moves towards the periphery in amplitude and shape. These changes may be largely due to the reflection of the pulse wave and the tapering down of the arteries towards the peripheral region. Pulse propagation in arteries is further complicated by frequency-dependent phase distortion. These phenomena have been described by O'Rourke and Gallagher (1996) and are discussed in the wider literature on pulse. The blood pressure pulse resembles the PPG blood-volume pulse, with similar changes occurring in vascular disease, such as damping and a loss of pulsatility. The damping has been associated with a reduction in vessel compliance and increased peripheral resistance, although these changes have yet to be fully explained (Allen, 2007). Likewise, PPG has been recognised and investigated as a tool for identifying vascular disease. Such diagnostic information extracted from PPG has the potential to be used for understanding the psychophysiological state of subjects responding to a system or the environment. The potential of the PPG pulse was reviewed by Murray and Foster (1996), who identified many characteristics of the PPG pulse (Figure 21.7).

### Pulse-Wave Analysis Technique

For pulse-wave analysis, manual measurement and feature-extraction techniques once were commonplace using media including chart-recorder paper and a ruler or photographic recording/magnetic tape (Sherebrin and Sherebrin, 1990). Recently, computing technology and software data analysis tools have made the sophisticated pre- and post-processing of physiological waveforms possible. MATLAB® (MathWorks, Inc.) is a good example of a digital signal processing tool for pulse-wave analysis. Because PPG measurements are sensitive to patient and/or probe-tissue movement artefacts, the automatic detection of a motion artefact and the extraction of good-quality signal are critical for precise analyses (Figure 21.8). Therefore, computer-based filtering, feature extraction and waveform averaging also have been employed in PPG pulse-wave analysis, including the analysis of frequency (Sherebrin and Sherebrin, 1990), joint time-frequency (Yan et al., 2005), artificial neural networks (Allen and Murray, 1999), systems identification and transfer function modelling (Millasseau et al., 2000), principal component analysis (Enríquez et al., 2002), non-linear and chaos theory (Bhattacharya et al., 2001), cross-correlation (Drinnan et al., 2001) and double-differentiation acceleration plethysmograms (Bortlotto et al., 2000).



**FIGURE 21.8** Examples of measurement artefact and the extremes in physiological variation that can be seen in PPG recordings. Each recording is from the index finger site over a period of 1 min and the artefact/physiological events are marked. (a) An episode of gross movement artefact or PPG probe cable shaken. (b) Hand or finger tremor, (c) coughing and (d) irregular breathing pattern such as a deep gasp or yawn. (From Allen, J., *Physiol. Measure.*, 28(3), 1, 2007.)

## SKIN CONDUCTANCE RESPONSE

### HISTORY

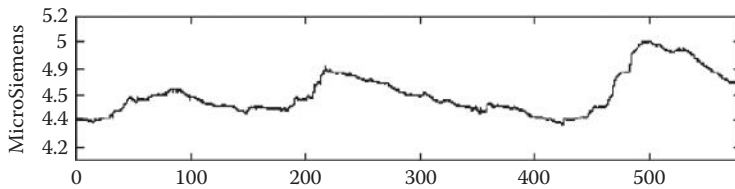
The measurement of conductance, also known as the SCR, galvanic skin response (GSR), electrodermal response (EDR) and psychogalvanic reflex (PGR), is a method of measuring the electrical conductance of the skin, which varies with its moisture level. This is interesting to observe because the sweat glands are controlled by the sympathetic nervous system (SNS) (Martini et al., 2003), so the increase of skin conductance is a detectable sign of psychological or physiological arousal.

The scientific study of GSR began in the early 1900s. One of the first references to the use of GSR instruments in psychoanalysis is in the book by C.G. Jung entitled *Studies in Word Analysis*, published in 1906. GSR was used in a variety of research fields in the 1960s through the late 1970s. As more sophisticated techniques such as EEG and MRI were developed, the use of GSR declined. However, skin conductance monitoring equipment is still in use because of its convenience and low cost.

### INSTRUMENTATION

The device measures the electrical conductance between two points; it is basically an ohm-meter measuring the resistance of the skin. Examples of SCR signals are shown in Figure 21.9. The two paths for current are along the surface of the skin and through the body. The conductance can be measured to several microsiemens ( $\mu\text{S}$ ). Because  $1 \text{ siemen (S)} = 1/\text{ohm } (\Omega) = 1 \text{ ampere (A)}/1 \text{ volt (V)}$ , the device can measure subtle changes in resistance on the skin and muscle tissue.





**FIGURE 21.9** A sample SCR signal during 600 s. (From Healey, J., Retrieval 11 July 2012, <http://vismod.media.mit.edu/tech-reports/TR-483/node6.html>, 1999.)

## ANALYSIS AND INTERPRETATION

A person's emotional arousal is controlled by sympathetic activity, although we cannot identify the specific emotion associated with the arousal. Nevertheless, SCR is a sensitive measure for detecting emotional responses, including fear, anger, the startle response, orienting responses and sexual feelings.

Bechara et al. (1994) and Bechara et al. (1999) investigated the SCR responses to questions related to decision-making in a gambling task. They reported that SCR exhibited not only the emotional outcomes of participants' choices but also the emotional arousal before their choice. This can be called as an anticipatory SCR which is the index of emotional arousal prior to their choice. That is, these anticipatory SCRs could be the indicator of whether the participants are about to make an advantageous or disadvantageous choice.

There is another interesting study on the subjective differences in SCRs. Osumi and Ohira (2010) observed a criminal investigation using a lie detector. The criminal, who was known to be a psychopath, did not show any variation in the SCR when fair and unfair offers were given as opposite stimulus, whereas normal healthy subjects showed significant differences in the SCR when the fair and unfair offers were given. This indicates that the SCR result of a lie detector test needs to be interpreted carefully based not only on the SCR but also a psychological diagnosis by a medical expert to avoid providing a psychopath with a legal advantage, especially in a criminal investigation.

## SKIN TEMPERATURE

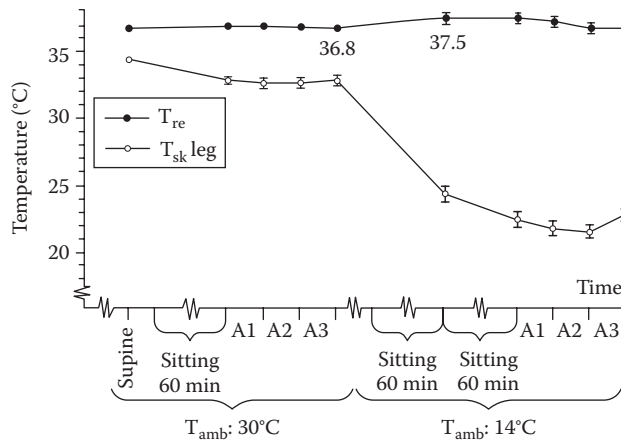
### THEORETICAL BACKGROUND

The SKT represents the outer surface temperature of the body. SKT is known to be related to blood flow in skin tissue as measured by plethysmography (Sargent et al., 1972). Thus, vasodilation and vasoconstriction of blood vessels can increase or decrease the SKT. An example of SKT signals is shown in Figure 21.10. Because the vasoconstriction is controlled by the SNS, which is a part of the autonomous nervous system, SKT provides an indirect measure of SNS activity. That is, SKT can be used to evaluate how much the SNS is activated by a particular environmental stimulus. Moreover, the SKT measurement can be performed at relatively low cost and the application of sensors on the fingers or toes is relatively convenient compared to other psychophysiological measures. Therefore, SKT is a very user-friendly tool for measuring SNS activity involving stress and relaxation.

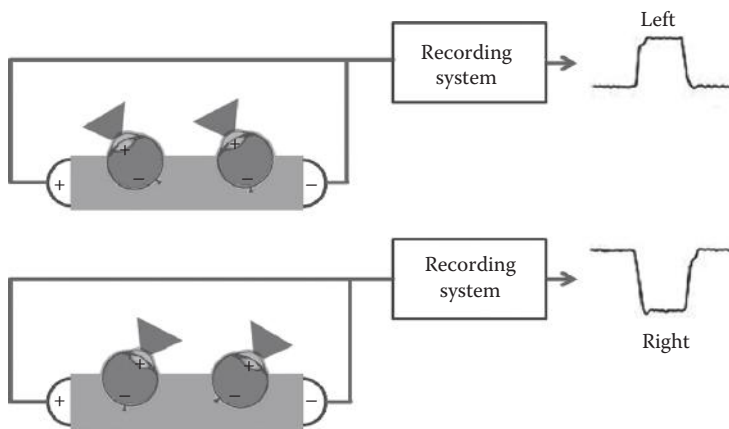
## ELECTROOCULOGRAPHY

### THEORETICAL BACKGROUND

The EOG technique measures the changes in electrical potential that occur when the eyes move, also known as the corneal–retinal potential. This potential is the result of the difference in potential between the front and the back of the eyeball. The potential of the cornea is 0.40–1.0 mV



**FIGURE 21.10** Temporal changes in average SKT ( $T_{sk}$ ) above the soleus muscle and rectal temperature ( $T_{re}$ ). A1 to A3 indicate periods of muscle activity. (From Winkel, J. and Jorgensen, K., *Euro. J. Appl. Physiol. Occup. Physiol.*, 63[5], 345, 1991.)



**FIGURE 21.11** Electrical charges of the eyeball.

positive with respect to the retina. The eye is a dipole in which the front pole is positive and the back is negative. This relationship is shown in Figure 21.11.

The EOG has been useful in a wide range of applications from the rapid-eye movements measured in sleep studies to the recording of visual fixations during perception, visual search and the experience of illusions in psychopathology. Reading, eye movements during real and simulated car driving, radar scanning and reading instrument dials under vibrating conditions are among the practical tasks examined with eye movement recordings (Andreassi, 2000).

## RECENT STUDIES AND ISSUES USING PSYCHOPHYSIOLOGICAL MEASURES

### BRAIN-COMPUTER INTERFACE

A Brain-Computer Interface (BCI) is literally a direct technological interface between a brain and a computer, without motor output from the user (Warwick, 2011). It is a combination of readily available computer hardware, a sophisticated adaptive algorithm and new physiological knowledge in a system that allows users to communicate through EEG activity.

A BCI translates the electrical activity from the body into commands used for controlling external devices such as a cursor on a computer screen (Wolpaw et al., 2002; Whang et al., 2003). In particular, a BCI using EEG signals is often used for controlling external devices such as spelling applications, wheelchairs and games. Edlinger et al. (2009, 2011) and Guger et al. (2008) designed a smart home that could be controlled by brainwaves. Postelnicu et al. (2012) created a set of seven masks – TV, phone, music, light, temperature, go to and move – and evaluated a hybrid neurophysiological interface for controlling a virtual smart home. EOG and EEG techniques have been used for recording biosignals from the bodies of participants. The signals were translated into commands for controlling a virtual smart home. Hwang et al. (2012) studied an effective and practical paradigm for a BCI-based 2D virtual wheelchair control. Two healthy female and three healthy male subjects, aged 23–30 years old, participated in this study. In the subjects, EEGs were recorded from 27 surface electrodes (Jasper and Andrews, 1938); with these EEG signals, subjects could operate a wheelchair and instruct it to turn left or turn right, go straight and stop.

### **AUTOMOBILE SAFETY**

Psychophysiologicalists have attempted to identify patterns of physiological signals during driving for the purpose of improving automobile safety. Karel and Waard (1993) and Gundel et al. (1995) demonstrated the possibility of using drivers' psychophysiological responses to understand their behaviour when driving. EEG data have been used to indicate alertness, sleepiness and workload when evaluating the mental state of drivers (Kecklund and Åkerstedt, 1993). Richter et al. (1998) reported that the eye blink rate decreased when driving on a complicated section of road. HR and HRV also have been shown to be sensitive to changes in a driver's mental workload, road conditions and driving duration (de Waard et al., 1995; Gobel et al., 1998; Richter et al., 1998). These studies were performed to evaluate driving behaviour and to use physiology to assess mental effort or stress. Recently, psychophysiologicalists have collected physiological measurements during actual, not simulated, driving conditions. This allows for a realistic interpretation of the driver's state during driving. Ramon et al. (2008) developed an integrated platform to assess a driver's physiological and functional states. Researchers have suggested the design and implementation of a platform that could measure physiological signals such as HR, respiration and skin resistance that are relevant indicators of a driver's mental state. Rigas et al. (2012) suggested a real-time methodology for the detection of stress events during real driving conditions. They used EKG, electrodermal activity and respiration to detect stress events. Kim et al. (2004a–c) developed and validated the model to diagnose the road–driver compatibility using biosignals such as EEG, SCR, SKT and HR and suggested a system design to enhance the driver's emotional satisfaction as well as to monitor the driver's safety and health condition. The purpose of these studies was to identify the relationship between descriptive statistics of psychophysiological data and driving behaviour during simulator or real driving conditions. However, variability in the psychophysiological signal often made it difficult to interpret results that were computed using a parametric approach. Therefore, to further interpret the meaning of the psychophysiological response of drivers, a variety of approaches should be developed to accurately evaluate and predict a driver's mental and physical state during driving.

### **MOTOR CONTROL**

Motor control involves the activity of the central nervous system (CNS) using information processing and the musculoskeletal system to create coordinated movements and skilled actions. Motor control is closely related with posture stability, which often influences slips and falls. Thus, researchers have attempted to measure posture stability using EMG, force plates and measurements of insole pressure. In particular, EMG is commonly used to measure the muscle activity and muscular co-contraction in posture stability research.

Parnianpour et al. (1988) investigated the effects of isodynamic fatiguing of flexion and extension trunk movements on the movement patterns and the motor output of the trunk. They observed that motor control of fatigued muscles was significantly lower and the range of motion was significantly greater in the coronal and transverse planes when subjects performed the primary task of flexion and extension. Granata and Wilson (2001) investigated the relationship between the trunk posture and musculoskeletal stability of the spine. They used EMG to observe the muscle recruitment patterns that maintain stability in various trunk postures. Brown and Potvin (2005) suggested an optimisation model of the spine. They reported that the inclusion of stability constraints in optimisation simulations produced realistic predictions of antagonist muscle activity and predictions of spine compression levels that agree more closely with EMG-based estimates compared to simulations in which stability was unconstrained. In construction workers working at two different heights on a movable scaffold frame, Min et al. (2011a,b) investigated postural stability by observing the spinal stability with trunk muscle recruitment, as well as cardiovascular stress and the workers' subjective perception of the difficulty of maintaining posture balance. Davarian et al. (2012) showed that EMG can be used to examine the clinical instability of chronic low-back patients that can be treated through a strength training protocol targeting trunk extensor muscles. Nugent et al. (2012) examined compartmentalisation in human lumbar spine extensor muscles during simple spine bending and twisting tasks to demonstrate independent neuromuscular control of the adjacent muscle of the lumbar erector spinae. They showed a difference in the timing of muscle activation above and below the third lumbar vertebra (L3) and a tendency towards pre-activation in the multifidus muscle compared to the erector spinae muscle during side bending. They suggested that these results have implications for understanding trunk control in the context of posture stability and posture adaptation during locomotion, as well as for lower back functional assessment and rehabilitation.

## CONCLUSIONS

Psychophysiological measurement has contributed in expanding the area of quantification of human psychology. During recent decades, along with the growing availability of new instruments for recording physiological responses, the number of psychophysiological studies has greatly increased to form an important part of psychology and neuroscience. We all know that interpreting the psychophysiological data is very challenging due to its great range of variability, which can result in even the most powerful form of statistical analysis finding very little of observed phenomena. However, data-handling challenges are being addressed by the development of sophisticated signal processing algorithms as well as advanced parametric and non-parametric data analysis schemes. Therefore, it is foreseen that the continuing development of methodology to measure both mental and physical workload could provide a solid mathematical ground to better resolve the human workload problems affecting various aspects of our working and living environments.

## REFERENCES

- Akasaki, K., K. Mita, M. Watakabe and K. Itoh (2001). Mechanomyogram and force relationship during voluntary isometric ramp contractions of the biceps brachii muscle. *European Journal of Applied Physiology*, 84, 19–25.
- Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), 1–29.
- Allen, J. and A. Murray (1999). Modeling the relationship between peripheral blood pressure and blood volume pulses using linear and neural network system identification techniques. *Physiological Measurement*, 20, 287–301.
- Allen, J. and A. Murray (2000). Similarity in bilateral photoplethysmographic peripheral pulse wave characteristics at the ears, thumbs and toes. *Physiological Measurement*, 21(3), 369–380.
- Allison, G.T., P. Godfrey and G. Robinson (1998). EMG signal amplitude assessment during abdominal bracing and hollowing. *Journal of Electromyography Kinesiology*, 8(1), 51–57.

- Andreassi, J.L. (2000). *Psychophysiology: Human Behaviour and Physiological Response*. Lawrence Erlbaum, Hillsdale, NJ.
- Balestra, G., M. Knaflitz and R. Merletti (1988). Comparison between myoelectric signal mean and median frequency estimates. *Proceedings of the Annual International Conference of the IEEE*, Seattle, WA.
- Basmajian, J.V. and R. Blumenstein (1980). *Electrode Placement in EMG Biofeedback*. Williams & Wilkins, Baltimore, MD.
- Basmajian, V. and C.J. Luca (1985). *Muscle Alive: Their Functions Revealed by Electromyography* (5th ed.). Williams & Wilkins, Baltimore, MD.
- Bechara, A., A.R. Damasio, H. Damasio and S.W. Anderson (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50, 7–15.
- Bechara, A., H. Damasio, A.R. Damasio and G.P. Lee (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of Neuroscience*, 19, 5473–5481.
- Beck, T.W., T.J. Housh, G.O. Johnson, J.P. Weir, J.T. Cramer, J.W. Coburn and M.H. Malek (2004). Mechanomyographic and electromyographic time and frequency domain responses during submaximal to maximal isokinetic muscle actions of the biceps brachii. *European Journal of Applied Physiology*, 92(3), 352–359.
- Beck, T.W., T.J. Housh, G.O. Johnson, J.P. Weir, J.T. Cramer, J.W. Coburn and M.H. Malek (2005). Comparison of Fourier and wavelet transform procedures for examining the mechanomyographic and electromyographic frequency domain responses during fatiguing isokinetic muscle actions of the biceps brachii. *Journal of Electromyography and Kinesiology*, 15(2), 190–199.
- Beck, T.W., T.J. Housh, J.P. Weir, J.T. Cramer, V. Vardaxis, G.O. Johnson, J.W. Coburn, M.H. Malek and M. Mielke (2006). An examination of the Runs Test, Reverse Arrangements Test, and modified Reverse Arrangements Test for assessing surface EMG signal stationarity. *Journal of Neuroscience Methods*, 156(1), 242–248.
- Berntson, G.G. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, 34, 623–648.
- Bhattacharya, J., P.P. Kanjilal and V. Muralidhar (2001). Analysis and characterization of photo-plethysmographic signal. *Biomedical Engineering, IEEE Transactions on*, 48(1), 5–11.
- Bilodeau, M., M. Cincera, A.B. Arsenault and D. Gravel (1997). Normality and stationarity of EMG signals of elbow flexor muscles during ramp and step isometric contractions. *Journal of Electromyography and Kinesiology*, 7(2), 87–96.
- Blanco, S., H. Garcia, R.Q. Quiroga, L. Romanelli and O. Rosso (1995). Stationarity of the EEG series. *Engineering in Medicine and Biology Magazine, IEEE*, 14(4), 395–399.
- Blinowska, K. and P. Durka (2006). *Electroencephalography (EEG)*, *Encyclopedia of Biomedical Engineering*. John Wiley & Sons, Inc., Hoboken, NJ.
- Bonato, P., G. Gagliati and M. Knaflitz (1996). Analysis of myoelectric signals recorded during dynamic contractions. *IEEE Engineering in Medicine and Biology*, 15, 102–111.
- Bortolotto, L.A., J. Blacher, T. Kondo, K. Takazawa and M.E. Safar (2000). Assessment of vascular aging and atherosclerosis in hypertensive subjects: Second derivative of photoplethysmogram versus pulse wave velocity. *American Journal of Hypertension*, 13, 165–171.
- Broman, H., G. Bilotto and C. De Luca (1985). Myoelectric signal conduction velocity and spectral parameters: Influence of force and time. *Journal of Applied Physiology*, 58(5), 1428–1437.
- Brown, S.H.M. and J.R. Potvin (2005). Constraining spine stability levels in an optimization model leads to the prediction of trunk muscle cocontraction and improved spine compression force estimates. *Journal of Biomechanics*, 38(4), 745–754.
- Cho, Y.J. and J.Y. Kim (2012). The effects of load, flexion, twisting and window size on the stationarity of trunk muscle EMG signals. *International Journal of Industrial Ergonomics*, 42(3), 287–292.
- Coburn, J.W., T.J. Housh, J.T. Cramer, J.P. Weir, J.M. Miller, T.W. Beck, M.H. Malek and G.O. Johnson (2005). Mechanomyographic and electromyographic responses of the vastus medialis muscle during isometric and concentric muscle actions. *The Journal of Strength and Conditioning Research*, 19(2), 412–420.
- Constant, I. and N. Sabourdin (2012). The EEG signal: A window on the cortical brain activity. *Pediatric Anesthesia*, 22(6), 539–552.
- Coorevits, P., L. Danneels, D. Cambier, H. Ramon, H. Druyts, K.J. Stefan, G.D. Moor and G. Vanderstraeten (2008). Correlations between short-time Fourier-and continuous wavelet transforms in the analysis of localized back and hip muscle fatigue during isometric contractions. *Journal of Electromyography and Kinesiology: Official Journal of the International Society of Electrophysiological Kinesiology*, 18(4), 637.
- Coumel, P., P. Maison-Blanche and D. Catuli (1995). Heart rate and heart rate variability. In: Malik, M. and Camm, A.J., (eds.), *Heart Rate Variability*. Futura, Armonk, NY, pp. 207–221.



- Criswell, E. (2011). *Cram's Introduction to Surface Electromyography*. Jones & Bartlett Publishers, Sudbury, MA.
- Dankaerts, W., P.B. O'Sullivan, A.F. Burnett, L.M. Straker and L.A. Danneels (2004). Reliability of EMG measurements for trunk muscles during maximal and sub-maximal voluntary isometric contractions in healthy controls and CLBP patients. *Journal of Electromyography Kinesiology*, 14, 333–342.
- Davarian, S., N. Maroufi, I. Ebrahimi, F. Farahmand and M. Parnianpour (2012). Trunk muscles strength and endurance in chronic low back pain patients with and without clinical instability. *Journal of Back and Musculoskeletal Rehabilitation*, 25(2), 123–129.
- Davidson, R.J. and N.A. Fox (1982). Asymmetrical brain activity discriminates between positive and negative stimuli in human infants. *Science*, 218(4578), 1235–1237.
- Davis, J.F. (1959). *Manual of Surface Electromyography*. Aerospace Medical Laboratory, Wright-Patterson Air Force Base, OH.
- de Waard, D., M. Jessurun, F.J.J.M. Steyvers, P.T.F. Reggatt and A.B. Karel (1995). Effect of road layout and road environment on driving performance, drivers' physiology and road appreciation. *Ergonomics*, 38(7), 1395–1407.
- Drinnan, M.J., J. Allen and A. Murray (2001). Relation between heart rate and pulse transit time during paced respiration. *Physiological Measurement*, 22(3), 425–432.
- Edlinger, G., C. Holzner, C. Groenegrass, C. Guger and M. Slater (2009). Goal-oriented control with brain-computer interface. Foundations of augmented cognition. *Neuroergonomics and Operational Neuroscience*, 5638, 732–740.
- Edlinger, G., C. Holzner and C. Guger (2011). A hybrid brain-computer interface for smart home control. Human-Computer Interaction. *Interaction Techniques and Environments*, 6762, 417–426.
- Eleanor, C. (2011). *Cram's Introduction to Surface Electromyography* (2nd ed.). Jone and Bartlett publishers, Boston, MA.
- Elul, R. (1972). The genesis of the EEG. *International Review of Neurobiology*, 15, 227–272.
- Enr'iquez, R.H., M.S. Castellanos, J.F. Rodríguez and J.L.H. C'aceres (2002). Analysis of the photoplethysmographic signal by means of the decomposition in principal components. *Physiological Measurement*, 23, N17–N29.
- Farina, D. and R. Merletti (2000). Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions. *Journal of Electromyography and Kinesiology: Official Journal of the International Society of Electrophysiological Kinesiology*, 10(5), 337–349.
- Forvi, E., M. Bedoni, R. Carabalona, M. Soncini, P. Mazzoleni, F. Rizzo, C. O'Mahony, C. Morasso, D.G. Cassarà and F. Gramatica (2012). Preliminary technological assessment of microneedles-based dry electrodes for biopotential monitoring in clinical examinations. *Sensors and Actuators A: Physical*, 180, 177–186.
- Fox, N.A. (1991). If it's not left, it's right: Electroencephalograph asymmetry and the development of emotion. *American Psychologist*, 46(8), 863–872.
- Gargiulo, G.D., P. Bifulco, M. Cesarelli, A. Fratini and M. Romano (2014). Problems in assessment of novel biopotential front-end with dry electrode: A brief review. *Machines*, 2(1), 87–98.
- Gerdle, B., S. Karlsson, A.G. Crenshaw, J. Elert and J. Fridén (2000). The influences of muscle fibre proportions and areas upon EMG during maximal dynamic knee extensions. *European Journal of Applied Physiology*, 81(1), 2–10.
- Gheab, N.H. and S.N. Saleem (2008). Comparison study of electromyography using wavelet and neural network. *Al-Khwarizmi Engineering Journal*, 4(3), 108–119.
- Gobel, M., J. Springer and J. Scherff (1998). Stress and strain of short haul bus drivers: Psychophysiology as a design oriented method for analysis. *Ergonomics*, 41(5), 563–580.
- Granata, K. and S. Wilson (2001). Trunk posture and spinal stability. *Clinical Biomechanics*, 16(8), 650–659.
- Guger, C., C. Holzner, C. Groenegrass, G. Edlinger and M. Slater (2008). Control of a smart home with a brain-computer interface. *Fourth International Brain-Computer Interface Workshop and Training Course 2008*, University of Technology Graz, Graz, Austria.
- Gundel, A., J. Drescher, H. Maas, A. Samel and M. Vejvoda (1995). Sleepiness of civil airline pilots during two consecutive night flights of extended duration. *Biological Psychology*, 40(1), 131–141.
- Hanowell, L., J.H. Eisele and D. Downs (1987). Ambient light affects pulse oximeters. *Anesthesiology*, 67(5), 864.
- Healey, J. (1999). Skin Conductance, Retrieval 11 July 2012, <http://vismod.media.mit.edu/tech-reports/TR-483/node6.html>.
- Hertzman, A.B. and C.R. Spealman (1937). Observation of the finger volume pulse recorded photoelectrically. *American Journal of Physiology*, 119, 334–335.
- Herwig, U., P. Satrapi and C. Schönfeldt-Lecuona (2003). Using the international 10–20 EEG system for positioning of transcranial magnetic stimulation. *Brain Topography*, 16(2), 95–99.



- Hoff, H.E. (1996). Galvani and the pregalvanian electrophysiologists. *Annals of Science*, 1, 157–172.
- Horita, T. and T. Ishiko (1987). Relationships between muscle lactate accumulation and surface EMG activities during isokinetic contractions in man. *European Journal of Applied Physiology and Occupational Physiology*, 56(1), 18–23.
- Hostens, I., J. Seghers, A. Spaepen and H. Ramon (2004). Validation of the wavelet spectral estimation technique in biceps brachii and brachioradialis fatigue assessment during prolonged low-level static and dynamic contractions. *Journal of Electromyography and Kinesiology*, 14, 205–215.
- Hwang, H.J., J.H. Lim, Y.J. Jung, H. Choi, S.W. Lee and C.H. Im (2012). Development of an SSVEP-based BCI spelling system adopting a QWERTY-style LED keyboard. *Journal of Neuroscience Methods*, 208(1), 59–65.
- Inbar, G.F. and A.E. Noujaim (1984). On surface EMG spectral characterization and its application to diagnostic classification. *Biomedical Engineering, IEEE Transactions on*, 9, 597–604.
- Jasper, H.H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371–375.
- Jasper, H.H. and H.L. Andrews (1938). Electroencephalography. III. Normal differentiation of occipital and precentral regions in man. *Archives of Neurology & Psychiatry (Chicago)*, 39, 96–115.
- Jönsson, P. (2007). Respiratory sinus arrhythmia as a function of state anxiety in healthy individuals. *International Journal of Psychophysiology*, 63, 48–54.
- Kanayama, N., A. Sato and H. Ohira (2007). Crossmodal effect with rubber hand illusion and gamma band activity. *Psychophysiology*, 44(3), 392–402.
- Karel, A.B. and D. de Waard (1993). The use of psychophysiology to assess driver status. *Ergonomics*, 36(9), 1099–1110.
- Karlsson, S., J. Yu and M. Akay (2000). Time-frequency analysis of myoelectric signals during dynamic contractions: A comparative study. *IEEE Transactions on Bio-Medical Engineering*, 47, 228–238.
- Kecklund, G. and T. Åkerstedt (1993). Sleepiness in long distance truck driving: An ambulatory EEG study of night driving. *Ergonomics*, 36, 1007–1017.
- Kim, J.Y., M.C. Jung and J.M. Haight (2005). The sensitivity of autoregressive model coefficient in quantification of trunk muscle fatigue during a sustained isometric contraction. *International Journal of Industrial Ergonomics*, 35, 321–330.
- Kim, J.Y. and S.Y. Yoon (2004a). The study to diagnose the road-driver compatibility I: Comparison of methods for bio-signal analysis. *Journal of the Korean Institute of Industrial Engineers*, 30(1), 44–49.
- Kim, J.Y. and S.Y. Yoon (2004b). The study to diagnose the road-driver compatibility III: Development and validation of diagnostic model. *Journal of the Korean Institute of Industrial Engineers* 30(1), 58–64.
- Kim, J.Y., S.Y. Yoon and J.S. Park (2004c). The study to diagnose the road-driver compatibility II: Data collection, variable selection and parameter quantification. *Journal of the Korean Institute of Industrial Engineers*, 30(1), 50–57.
- Kisley, M.A. and Z.M. Cornwell (2006). Gamma and beta neural activity evoked during a sensory gating paradigm: Effects of auditory, somatosensory and cross-modal stimulation. *Clinical Neurophysiology*, 117(11), 2549–2563.
- Klados, M.A., C. Styliadis and P.D. Bamidis. (2014). A short review on emotional recognition based on bio-signal pattern analysis. *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, Seville, Spain, pp. 787–790.
- Knaflitz, M. and P. Bonato (1999). Time-frequency methods applied to muscle fatigue assessment during dynamic contractions. *Journal of Electromyography and Kinesiology*, 9(5), 337–350.
- Lindström, L., R. Magnusson and I. Petersen (1970). Muscular fatigue and action potential conduction velocity changes studied with frequency analysis of EMG signals. *Electromyography*, 10(4), 341–356.
- Lo Conte, L. and R. Merletti (1995). Advances in processing of surface myoelectric signals: Part 2. *Medical and Biological Engineering and Computing*, 33(3), 373–384.
- MacIsaac, D., P.A. Parker and R.N. Scott (2001). The short-time Fourier transform and muscle fatigue assessment in dynamic contractions. *Journal of Electromyography and Kinesiology*, 11, 439–449.
- Marras, W.S. and K.G. Davis (2001). A non-MVC EMG normalization technique for the trunk musculature: Part 1. Method development. *Journal of Electromyography and Kinesiology*, 11, 1–9.
- Marras, W.S., K.G. Davis, C.A. Heaney, A.B. Maronitis and W.G. Allread (2000). The influence of psychosocial stress, gender, and personality on mechanical loading of the lumbar spine. *Spine*, 25(23), 3045–3054.
- Marras, W.S., S.L. Rangarajulu and S.A. Lavender (1987). Trunk loading and expectation. *Ergonomics*, 30, 551–562.
- Martini, F. and E. Bartholomew (2003). *Essentials of Anatomy & Physiology*. Benjamin Cummings, San Francisco, CA.

- Matthes, K. (1935). Untersuchungen über die Sauerstoffsättigung des menschlichen Arterienblutes. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 179(6), 698–711.
- Merletti, R., M. Knaflitz and C. De Luca (1992). Electrically evoked myoelectric signals. *Critical Reviews in Biomedical Engineering*, 19(4), 293–340.
- Merletti, R. and L.R. Lo Conte (1997). Surface EMG signal processing during isometric contractions. *Journal of Electromyography and Kinesiology*, 7(4), 241–250.
- Mesulam, M.M. and J. Perry (1972). The diagnosis of love-sickness: Experimental psychophysiology without the polygraph. *Psychophysiology*, 9(5), 546–551.
- Millasseau, S.C., F.G. Guigui, R.P. Kelly, K. Prasad, J.R. Cockcroft, J.M. Ritter and P.J. Chowienczyk (2000). Noninvasive assessment of the digital volume pulse comparison with the peripheral pressure pulse. *Hypertension*, 36(6), 952–956.
- Millet, D. (2002). The Origins of EEG. *Seventh Annual Meeting of the International Society for the History of the Neurosciences (ISHN)*, Los Angeles, CA.
- Min, S.N., J.Y. Kim and M. Parnianpour (2011a). Effects of construction worker's experience, the presence of safety handrail and height of movable scaffold on postural and spinal stability. *First Middle East Conference on Biomedical Engineering (MECBME), 2011*, Sharjah, United Arab Emirates, IEEE, pp. 146–149.
- Min, S.N., J.Y. Kim and M. Parnianpour (2011b). The effects of safety handrails and the heights of scaffolds on the subjective and objective evaluation of postural stability and cardiovascular stress in novice and expert construction workers. *Applied Ergonomics*, 43(3), 574–581.
- Murray, W.B. and P.A. Foster (1996). The peripheral pulse wave: Information overlooked. *Journal of Clinical Monitoring*, 12, 365–377.
- Nickel, P. and F. Nachreiner (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors*, 45(4), 575–590.
- Nimbarte, A.D., M.J. Al Hassan, S.E. Guffey and W.R. Myers (2012). Influence of psychosocial stress and personality type on the biomechanical loading of neck and shoulder muscles. *International Journal of Industrial Ergonomics*, 42(5), 397–405.
- Nugent, M.M., P.J. Stapley and T.E. Milner (2012). Independent activation in adjacent lumbar extensor muscle compartments. *Journal of Electromyography and Kinesiology*, 22(4), 531–539.
- Oppenheim, A.V., R.W. Schaffer and J.R. Buck (1989). *Discrete-Time Signal Processing*. Prentice hall Englewood Cliffs, Prentice Hall, NJ.
- O'Rourke, M.F. and D.E. Gallagher (1996). Pulse wave analysis. *Journal of Hypertension*, 14(5), 147–157.
- Osumi, T. and H. Ohira (2010). The positive side of psychopathy: Emotional detachment in psychopathy and rational decision-making in the ultimatum game. *Personality and Individual Differences*, 49(5), 451–456.
- Paiss, O. and G.F. Inbar (1987). Autoregressive modeling of surface EMG and its spectrum with application to fatigue. *Biomedical Engineering, IEEE Transactions on*, 10, 761–770.
- Parnianpour, M., M. Nordin, N. Kahanovitz and V. Frankel (1988). 1988 Volvo award in biomechanics. The tri-axial coupling of torque generation of trunk muscles during isometric exertions and the effect of fatiguing isoinertial movements on the motor output and movement patterns. *Spine*, 13(9), 982–992.
- Perry, S.R., T.J. Housh, J.P. Weir, G.O. Johnson, A.J. Bull and K.T. Ebersole (2001). Mean power frequency and amplitude of the mechanomyographic and electromyographic signals during incremental cycle ergometry. *Journal of Electromyography and Kinesiology: Official Journal of the International Society of Electrophysiological Kinesiology*, 11(4), 299–305.
- Petrofsky, J. (1979). Frequency and amplitude analysis of the EMG during exercise on the bicycle ergometer. *European Journal of Applied Physiology and Occupational Physiology*, 41(1), 1–15.
- Phillips, R. and M. Feeney (1973). *The Cardiac Rhythms: A Systematic Approach to Interpretation*. W. B. Saunders Co, Philadelphia, PA.
- Popivanov, D. and A. Todorov (1986). Statistical procedures for interference EMG power spectra estimation. *Medical and Biological Engineering and Computing*, 24(4), 344–350.
- Postelnicu, C.C., A. Covaci, A. Panfir and D. Talaba (2012). Evaluation of a P300-based interface for smart home control. *Technological Innovation for Value Creation*, 372, 179–186.
- Ramon C., A. Clarion, C. Gehin, C. Petit, C. Collet and A. Dittmar (2008). An integrated platform to assess driver's physiological and functional states. *30th Annual International Conference on the IEEE*, Piscataway, NJ, Engineering in Medicine and Biology Society, pp. 506–509.
- Richter, P., T. Wagner, R. Heger and G. Weise (1998). Psychophysiological analysis of mental load during driving on rural roads—a quasi-experimental field study. *Ergonomics*, 41(5), 593–609.
- Rigas, G., Y. Goletsis and D.I. Fotiadis (2012). Real-time driver's stress event detection. *Intelligent Transportation Systems, IEEE Transactions on*, 13(1), 221–234.

- Ronald M. Birse, rev. Patricia E. Knowlden (<http://0-www.oxforddnb.com/innopac.up.ac.za:80/view/article/37794>) Oxford Dictionary of National Biography 2004 (Subscription required) (Original source is his biography written by his wife – Elizabeth Muirhead. Alexander Muirhead 1848–1920. Oxford, Blackwell: privately printed 1926).
- Roy, S.H., P. Bonato and M. Knaflitz (1998). EMG assessment of back muscle function during cyclical lifting. *Journal of Electromyography and Kinesiology*, 8(4), 233–245.
- Sargent, J.D., E.E. Green and E. Walters (1972). The use of autogenic feedback training in a pilot study of migraine and tension headaches. *Headache: The Journal of Head and Face Pain*, 12(3), 120–124.
- Schleifer, L.M., T.W. Spalding, S.E. Kerick, J.R. Cram, R. Ley and B.D. Hatfield (2008). Mental stress and trapezius muscle activation under psychomotor challenge: A focus on EMG gaps during computer work. *Psychophysiology*, 45(3), 356–365.
- Schneider, S. and H.K. Struder (2012). EEG-theoretical background and practical aspects 2. In: Boecker, H. (ed.), *Imaging Mood Induction and Pain Processing in Exercise*. Springer, Heidelberg, Germany.
- Seroussi R.E. and M.H. Pope (1987). The relationship between trunk muscle electromyography and lifting moments in the sagittal and frontal planes. *Journal of Biomechanics*, 20, 135–146.
- Sherebrin, M. and R. Sherebrin (1990). Frequency analysis of the peripheral pulse wave detected in the finger with a photoplethysmograph. *Biomedical Engineering, IEEE Transactions on*, 37(3), 313–317.
- Solomonow, M., C. Baten, J. Smit, R. Baratta, H. Hermens, R. D'Ambrosia and H. Shoji (1990). Electromyogram power spectra frequencies associated with motor unit recruitment strategies. *Journal of Applied Physiology*, 68(3), 1177–1185.
- Stassen, H., D. Lykken and G. Bombent (1988). The within-pair EEG similarity of twins reared apart. *European Archives of Psychiatry and Clinical Neuroscience*, 237(4), 244–252.
- Stern, R.J., W.J. Ray and C.M. Davis (1980). *Psychophysiological Recording*. Oxford University Press, New York.
- Stern, R.M., W.J. Ray and K.S. Quigley (2001). *Psychophysiological Recording*. Oxford University Press, New York.
- Taheri, B.A., R.T. Knight and R.L. Smith (1990). A dry electrode for EEG recording. *Electroencephalography and Clinical Neurophysiology*, 90(5), 376–383.
- Tomarken, A.J., R.J. Davidson, R.E. Wheeler and L. Kinney (2007). Psychometric properties of resting anterior EEG asymmetry: Temporal stability and internal consistency. *Psychophysiology*, 29(5), 576–592.
- Torse, D., R. Maggavi and S. Pujari (2012). Nonlinear blind source separation for EEG signal pre-processing in brain-computer interface system for epilepsy. *International Journal of Computer Applications*, 50(14), 12–19.
- U.S. Department of Health and Human Service (1992). *Selected Topics in Surface Electromyography for Use in the Occupational Setting: Expert Perspective*. National Institute for Occupational Safety and Health, Washington DC.
- Waller, A.D. (1887). A demonstration on man of electromotive changes accompanying the heart's beat. *The Journal of Physiology*, 8(5), 229–234.
- Warwick, K. (2011). *Artificial Intelligence: The Basics*. Taylor & Francis Group, Retrieval 23 April 2011, <http://books.google.com/books?id=yLW7bwAACAAJ>.
- Webster, J.G. (1997). *Design of Pulse Oximeters*. Taylor & Francis Group, Madison Avenue, New York.
- Whang, M.C., J.S. Lim and W. Boucsein (2003). Preparing computers for affective communication: A psychophysiological concept and preliminary results. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(4), 623–634.
- Wheeler, K.R. and C.C. Jorgensen (2003). Gestures as input: Neuroelectric joysticks and keyboards. *IEEE Pervasive Computing*, 2(2), 56–61.
- Wieringa, F.P., F. Mastik and A.F. van der Steen (2005). Contactless multiple wavelength photoplethysmographic imaging: A first step toward “SpO2 camera” technology. *Annals of Biomedical Engineering*, 33(8), 1034–1041.
- Winkel J. and K. Jørgensen (1991). Significance of skin temperature changes in surface electromyography. *European Journal of Applied Physiology and Occupational Physiology*, 63(5), 345–348.
- Wolpaw, J.R., N. Birbaumer, D.J. McFarland, G. Pfurtscheller and T.M. Vaughan (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6), 767–791.
- Yan Y.S., C.C. Poon and Y.T. Zhang (2005). Reduction of motion artifact in pulse oximetry by smoothed pseudo Wigner–Ville distribution. *Journal of NeuroEngineering Rehabilitation*, 2(3), R1–R39.



---

# 22 Neuroergonomic Methods for the Evaluation of Physical and Cognitive Work

*Raja Parasuraman and Ranjana Mehta*

## CONTENTS

|                                                       |     |
|-------------------------------------------------------|-----|
| Introduction to Neuroergonomics .....                 | 609 |
| Neuroergonomic Methods.....                           | 610 |
| Magnetic Resonance Imaging .....                      | 611 |
| Diffusion Tensor Imaging.....                         | 612 |
| Functional Magnetic Resonance Imaging .....           | 612 |
| Functional Near-Infrared Spectroscopy .....           | 613 |
| Transcranial Doppler Sonography .....                 | 613 |
| Electroencephalography .....                          | 614 |
| Event-Related Potentials .....                        | 615 |
| Non-Invasive Brain Modulation.....                    | 615 |
| Critical Issues in Cognitive Neuroergonomics .....    | 616 |
| Assessment of Mental Workload.....                    | 616 |
| Vigilance and Mental Fatigue .....                    | 619 |
| Neuroadaptive Systems .....                           | 621 |
| Training .....                                        | 622 |
| Critical Issues in Physical Neuroergonomics.....      | 623 |
| Assessment of Physical Work .....                     | 624 |
| Force Production .....                                | 624 |
| Muscle Flexion and Extension .....                    | 626 |
| Static and Dynamic Exertions .....                    | 627 |
| Muscle Fatigue .....                                  | 627 |
| Implications for Health and Performance .....         | 628 |
| Interactions between Physical and Cognitive Work..... | 630 |
| Effects of Physical Work on Cognitive Function .....  | 630 |
| Effects of Cognitive Work on Physical Capacity .....  | 630 |
| Conclusions and Future Directions.....                | 632 |
| References.....                                       | 632 |

## INTRODUCTION TO NEUROERGONOMICS

Neuroergonomics shares the same goal as do the other ergonomic methods that are described in this book: to help develop technological systems and work environments compatible with human capabilities and limitations, but by taking *both* human performance and brain function into account. Neuroergonomics can therefore be defined as the study of the human brain in relation

to performance at work and in everyday settings, including both physical and cognitive work (Karwowski et al. 2003; Parasuraman 2003; Parasuraman and Rizzo 2007).

The phenomenal advancement in our understanding of the neural mechanisms of human cognitive, motor and affective functions spurred the development of the field of neuroergonomics and continues to fuel its growth. Neuroscience has provided not only new methods with which to study the brain at work, such as neuroimaging, but also theories and empirical findings on brain function that can constrain, refine or extend ergonomic theories of human performance (Posner 2012). Neuroergonomics can therefore provide added value, beyond that available from traditional neuroscience and conventional ergonomics, to our understanding of brain function and behaviour as encountered in work and in other natural settings (Parasuraman 2011).

This chapter discusses several examples that illustrate the value of the neuroergonomic approach to human factors and ergonomics. We first describe the various methods that can be used in neuroergonomic evaluations of human work, emphasizing measures of brain function, as opposed to peripheral autonomic nervous system measures that are described in Chapter 21. We then focus on the use of these methods to examine selected critical issues in physical and cognitive work, beginning with the former before going on to the latter, followed by a discussion of the interactions between the two. We end with a summary of the main findings and discuss future prospects and directions. Our focus on *both* physical and cognitive work in this chapter is consistent with recent calls in Ergonomics/Human Factors (E/HF) to *put the mind and body back together* (Marras and Hancock 2014).

## NEUROERGONOMIC METHODS

The methods used in neuroergonomics stem from developments in neuroscience and *cognitive* neuroscience in particular. The past two decades have witnessed a revolution in neuroscience. This field was once restricted to examining brain structure and function invasively in animals or occasionally in human patients undergoing neurosurgery. However, the development of non-invasive neuroimaging techniques, first with positron emission tomography (PET) and later with functional magnetic resonance imaging (fMRI), led to an explosion of research using these methods to examine cognitive, motor and affective functions in living, healthy humans. The resulting knowledge base has considerably enhanced our understanding of the neural bases of human behaviour.

Neuroimaging techniques applicable to neuroergonomics fall into two general categories: those that are based on cerebral haemodynamics, such as fMRI, and those that assess brain electromagnetic activity, such as electroencephalography (EEG) and event-related potentials (ERPs). In terms of real-time measurement of brain function, haemodynamic methods have low temporal resolution (on the order of about 10 s), since they are based on changes in cerebral blood flow, which take place at a relatively sluggish rate. Electromagnetic methods, on the other hand, track brain changes at a much faster rate, on the order of milliseconds. This advantage in temporal resolution is countered by the lower spatial resolution of electromagnetic measures (such as EEG) in comparison to some haemodynamic measures, particularly fMRI. The higher spatial resolution (on the order of a few millimetres) of fMRI allows researchers to identify specific brain regions associated with the performance of a cognitive or physical task, whereas this is more difficult to achieve with electromagnetic methods such as EEG or ERPs.

In addition to these neuroimaging measures, psychophysiological measures that are reflections of autonomic nervous system activity, such as heart rate, blood pressure and galvanic skin response, are also useful adjuncts that can be added to the toolbox of neuroergonomic methods. Other psychophysiological measures include eye blinks and eye movements. Finally, neuroimaging and psychophysiology are *sensing* or *recording* techniques that assess different aspects of central and autonomic nervous system function. Complementing these sensing methods are brain *modulation* methods, in which neural activity is increased or decreased by the application of an external magnetic or electrical field to the scalp. These include transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS).



TABLE 22.1

## List of Neuroergonomic Techniques and Their Major Features

| Method | Measures/Stimulates                              | Portability | Purchase Cost                         | Operational Cost            | Spatial Resolution | Temporal Resolution |
|--------|--------------------------------------------------|-------------|---------------------------------------|-----------------------------|--------------------|---------------------|
| MRI    | Grey matter volume                               | None        | Very high (~\$2 million)              | High (\$500–\$600 per hour) | High               | NA                  |
| DTI    | White matter integrity                           | None        | Very high (~\$2 million)              | High (\$500–\$600 per hour) | High               | NA                  |
| fMRI   | Relative blood oxygenation                       | None        | Very high (~\$2 million)              | High (\$500–\$600 per hour) | High               | Low                 |
| fNIRS  | Oxyhaemoglobin and deoxyhaemoglobin              | High        | Moderate to high (\$25,000–\$200,000) | Minimal                     | Moderate           | Low                 |
| TCD    | Cerebral blood flow velocity                     | Moderate    | Moderate (~\$25,000)                  | Minimal                     | Low                | Low                 |
| EEG    | Post-synaptic electrical activity                | Moderate    | Moderate to high (\$25,000–\$80,000)  | Minimal                     | Low                | High                |
| ERP    | Stimulus or response-related electrical activity | Moderate    | Moderate to high (\$25,000–\$80,000)  | Minimal                     | Low                | High                |
| TMS    | Brain activation or inhibition                   | Low         | High (~\$80,000)                      | Minimal                     | High               | High                |
| tDCS   | Brain activation or inhibition                   | High        | Very low (\$100)                      | Minimal                     | Low                | Low                 |

Table 22.1 lists these methods and their major characteristics, such as portability, acquisition and operational costs and spatial and temporal resolution. We describe each of these techniques briefly in turn before discussing their use for the assessment of physical and cognitive work.

## MAGNETIC RESONANCE IMAGING

Magnetic resonance imaging (MRI) has revolutionized medicine with its ability to provide radiologists and other physicians with detailed, high-resolution images of the living human brain. Because it is non-invasive, MRI has also been of immense value to researchers and clinicians in neuroscience by revealing the features of brain anatomical structures in both healthy and patient populations. Even small parts of the brain such as the hippocampus, which is a critical structure involved in human learning and memory, can be visualized to examine, for example, whether it is damaged due to a tumour or because of a degenerative condition like Alzheimer's disease. However, because MRI provides *static* images of the physical brain, the reader might wonder why it would be of value to those interested in investigating aspects of brain *function* in healthy individuals.

A major reason is that other dynamic measures of human brain function, such as fMRI (see later in this chapter), can be overlaid on the static MRI image so that areas of activation as seen in the fMRI scan can be associated with particular brain regions. This is the standard use of MRI by researchers in cognitive neuroscience who employ fMRI to investigate the neural bases of human cognitive and motor functions. Second, and perhaps more importantly, although the MRI provides a static, structural image of the brain at a particular point in time, the physical structure of the brain itself can change over time, particularly in response to novel environmental inputs. That is, the adult human brain has a degree of *plasticity*, so that new learning or training can change its

structure (Huttenlocher 2002). We describe studies of brain plasticity in a later part of this chapter devoted to training. A third reason for the usefulness of structural MRI lies in its ability to provide distinct images of both grey matter, which contains neuronal cell bodies, and white matter, which refers to the collection of axons that connect neurons with each other in the brain. (The previous imaging technology, computerized tomography or CT, could not separate grey and white matter, which is why MRI largely supplanted it.) Effects of training on brain structure have mainly focused on changes in cortical grey matter. However, recent studies have shown that white matter is also subject to plasticity (Scholz et al. 2009). A recently developed neuroimaging technique called diffusion tensor imaging (DTI) can allow for experience-related changes in white matter to be visualized and quantified.

### Diffusion Tensor Imaging

The DTI technique uses MRI to target the diffusion of water molecules in the axons that make up white matter. When placed in a free medium, water molecules diffuse in all directions equally; that is, the diffusion is random or isotropic. The axons of neurons have a given orientation so that water molecules move more in this direction compared to movements in other directions, that is, the diffusion within an axon is non-random or anisotropic. DTI tracks the relative diffusion of water molecules in several directions and allows for the computation of the fractional anisotropy (FA), or the proportional diffusion along the axonal direction relative to all other directions.

The cellular basis of anisotropic water diffusion in the brain is not fully understood, but probably reflects the ordered arrangement of axons as well as the degree of myelination of axons and their diameter (Sen and Basser 2005). FA values can range from 0 to 1, where 0 indicates non-directional (isotropic) and 1 indicates perfectly directional (anisotropic) diffusion. FA is a frequently used index of white matter integrity that quantifies the magnitude and directionality of water diffusion in the brain. Higher FA values are thought to reflect greater integrity of white matter, linking different cortical and subcortical regions of the brain. For example, the FA value of a particular white matter pathway may increase from 0.3 to 0.4 with extended training. We describe such training studies later in this chapter.

### Functional Magnetic Resonance Imaging

The revolution in cognitive neuroscience owes much to functional MRI, which has become the principle neuroimaging technique used in this field. Such a leading position had previously been held by PET. However, the disadvantages of PET – its use of ionizing radiation, which means that only a limited number of scans can be performed on a participant in a given period of time, and its lower spatial resolution – led to its replacement in cognitive neuroscience research by fMRI.

Both PET and fMRI provide indirect measures of brain activity by capitalizing on a fundamental physiological property of the brain: When neural activity in a brain region increases, oxygenated blood flows to that area. PET uses injected radioactive tracers to measure the blood flow, whereas fMRI uses natural *tracers* in the blood – oxyhaemoglobin (HbO<sub>2</sub>) and deoxyhaemoglobin (HbR) – to index relative oxygenation. Because more oxygenated blood is delivered to the brain region than is used by neurons, the net result is a decrease in HbR in that region. Furthermore, HbO<sub>2</sub> and HbR have different magnetic properties, so the decrease can be detected in the MRI signal. The resulting contrast between oxygenated and deoxygenated blood is called the blood oxygenation level-dependent (BOLD) signal (Poldrack et al. 2011).

Since its development in the mid-1990s, BOLD fMRI has become the workhorse of hundreds of studies of the neural bases of cognitive, motor and affective functions. Although the majority of fMRI studies have used relatively simple laboratory tasks, as the field has progressed, researchers have turned to examine brain activation patterns in increasingly complex tasks more representative of natural everyday situations. These include studies of spatial navigation (Maguire 2007), simulated driving (Calhoun and Pearlson 2012) and flight (Callan et al. 2012) and viewing of natural scenes such as city views (Spiers and Maguire 2007).

## FUNCTIONAL NEAR-INFRARED SPECTROSCOPY

Functional near-infrared spectroscopy (fNIRS) is a non-invasive optical technique for measuring cerebral haemodynamics similar to PET and fMRI but with lower spatial resolution. A typical fNIRS system uses infrared light directed to the scalp through emitters. The light travels a short distance through the brain and exits the scalp, where it is captured by detectors (optodes) embedded in the same sensor strip as the emitters. As the light travels through brain tissue, it undergoes multiple scattering and partial absorption. Just as oxygenated and deoxygenated blood can be contrasted by their respective magnetic characteristics (as picked up by fMRI), they also differ in their optical absorption properties. Spectroscopic analysis allows for the detection of HbO<sub>2</sub> and HbR based on the absorption rate. Neural activity is associated with the conversion of HbO<sub>2</sub> to HbR. This cerebral haemodynamic response can be captured by fNIRS. Thus, fNIRS can effectively monitor changes in cerebral oxygenation and blood volume (Ide and Secher 2000; Maikala 2010; Villringer and Chance 1997). Other parameters that can be obtained are total haemoglobin, computed as the sum of HbO<sub>2</sub> and HbR, which indicates the level of the local cerebral blood volume, and individual HbO<sub>2</sub> and HbR concentrations, which indicate the interaction between the regional blood volume, the blood flow and the metabolic rate of oxygen (Thomas and Stephane 2008).

fNIRS is safe, highly portable and inexpensive, at least in comparison to PET and MRI. It can be quickly applied to a participant's head, is well tolerated and involves no operational expense, as does fMRI (typically \$500–\$600 per hour). Task-related cortical activation as assessed with fNIRS also has good test–retest reliability (Plichta et al. 2006). Changes in cerebral haemodynamics in cognitive tasks have been examined in many studies using fNIRS. For example, the prefrontal cortex (PFC) is activated during working memory tasks, with increased activation for higher memory load (Schreppel et al. 2008).

fNIRS has also been used in physical ergonomics. Several studies have demonstrated concomitant increases in cerebral oxygenation with increased motor activity, reflecting the contributions of the descending central command to perform motor tasks (Shibuya and Kuboyama 2007; Thomas and Stephane 2008). Compared to other haemodynamic imaging techniques, such as fMRI and PET, fNIRS has been shown to be an effective tool in quantifying cortical activation during dynamic movements, without causing movement artefact issues (Perrey 2008).

## TRANSCRANIAL DOPPLER SONOGRAPHY

Another measure of cerebral haemodynamics that can be used in ergonomics is transcranial Doppler sonography (TCD). Whereas fNIRS uses infrared light, TCD uses ultrasound to image cerebral blood flow (Aaslid 1986). TCD draws on the *Doppler* effect, which refers to the change in the measured frequency of sound (or light) waves when the receiver and the transmitter move relative to another (e.g. the pitch of an ambulance siren appears to get higher as it approaches a pedestrian and then lower as it moves past). TCD uses an emitter attached to the head to direct ultrasound towards an artery within the brain. A receiver (typically enclosed with the emitter) then records the frequency of the sound wave reflected by red blood cells moving through the artery. The magnitude of the change in frequency (the Doppler shift) varies directly proportional to the velocity of blood flow within the artery (Duschek and Schandry 2003).

The TCD transducer unit is typically positioned over the middle temporal window of the skull to image the middle cerebral artery (MCA). The MCA is the largest carrier of blood flow within each hemisphere, so it is the most frequently imaged artery, although occasionally, the anterior and posterior cerebral arteries are also studied. Cerebral blood flow velocity from the MCA provides a global index within each hemisphere of haemodynamic changes associated with neural activation. The MCA and other arteries in the brain deliver blood in response to the changing metabolic demands of active neurons. When an area of the brain becomes metabolically active due to task-related neuronal activity, by-products of this activity, such as carbon dioxide, increase, leading to

an increase in blood flow to that region to remove the waste product (Aaslid 1986). Consequently, TCD allows for the measurement of cerebral haemodynamics during cognitive and physical task performance.

A number of studies have used TCD to examine cerebral haemodynamics of perceptual, cognitive and motor performance. Because blood flow changes can be measured simultaneously in both cerebral hemispheres, several researchers have used basic tasks to examine hemispheric differences in cortical activation, for example, between language (left hemisphere) and visuospatial (right hemisphere) tasks (see reviews by Duschek and Schandry 2003; Stroobant and Vingerhoets 2000). Studies more relevant to E/HF have used TCD to assess changes in cerebral haemodynamic associated with supervisory control and vigilance tasks (see later in this chapter).

TCD has also been used in studies of physical work. For example, Hellström and Wahlgren (1993) examined blood flow velocity in the MCA following moderate and hard physical exercise on a bicycle ergometer. They reported that moderate exercise increased MCA blood flow velocity, whereas subsequent hard exercise led to a decrease, possibly due to MCA constriction. Dickerman et al. (2000) investigated blood flow velocity in the MCA during maximal weight lifting (static exercise) and at rest. They found that blood flow velocities were significantly decreased in all subjects during maximal lifting compared to the rest condition. On the other hand, Ogoh et al. (2010) reported no changes in mean cerebral blood flow velocities between rest, static exercise (30% maximal handgrip for 30 s) and recovery conditions. The discrepancies regarding the influence of physical work on cerebral blood flow velocity may be attributed in part to varying exercise levels and types that differentially affect systemic changes such as heart rate and blood pressure (Ainslie et al. 2007).

## ELECTROENCEPHALOGRAPHY

EEG has a long history of use in E/HF research (Gevins and Smith 2007 and see Chapter 21). However, recent developments in its miniaturization and portability have enhanced its value. The EEG signal recorded at the scalp, at any given moment, is due to the spatial summation of current density induced by synchronized post-synaptic potentials occurring in large clusters of neurons (Pizzagalli 2007). The EEG is recorded as a difference in voltage between an active electrode attached to the scalp and a reference electrode, typically the ear. EEG can be recorded simultaneously from multiple electrodes at different positions on the scalp, such as the frontal, parietal, temporal and occipital lobes of the brain according to the International 10–20 system. Typically, researchers employ from 32 to 256 electrode sites embedded in an elastic cap or net and attached to the scalp, depending on the goals of the study. EEG signals are typically analysed using spectral methods to assess power in different frequency bands: delta (0.5–3 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz) and gamma (40–50 Hz). Activity in each of these frequency bands has been associated with various cognitive and physical states.

EEG provides a useful measure that can be used to assess brain function in complex tasks, is well tolerated by operators and is relatively unobtrusive so that it does not interfere with the operator's work performance. Its compact size and low cost, compared to other neuroimaging techniques such as fMRI, makes it fairly well suited for use in both laboratory and field conditions. In the latter case, however, much care must be taken to ensure proper artefact removal. Several algorithms have been developed to allow for the removal of noise in the EEG signal due to eye blinks, muscle activity, etc., which can be used for artefact removal in real time (Jung et al. 2000). In addition, recently *dry* electrode caps, which do not need extensive participant preparation time, as well as wireless systems that do not require the participant to be tethered to a cable, have been developed Makeig et al. (2009). Each of these technical developments has enhanced the relevance and value of EEG for ergonomic applications.

Several studies have used EEG to examine different issues in the ergonomics of physical work. For example, significant increase in EEG activity, across multiple electrode sites and across various

frequencies bands, has been associated with exercise (Bailey et al. 2008; Kubitz and Mott 1996) and physical work (Krause et al. 1983) that indicates increased contribution of central command.

Several measures have been derived from EEG that distinguish muscle activation in different types of motor activities: isometric force production, static versus dynamic conditions, concentric versus eccentric conditions, joint control and strength, and fatigue and recovery. Coherence is one such measure used to quantify the linear correlation between two signals in the frequency domain (Halliday et al. 1995). Corticomuscular coherence (CMC) is determined as the coherence between sensorimotor cortex activation obtained from EEG and muscular activation as measured by electromyography (EMG) during motor activities. CMC reflects communications between the brain and the muscle, which is considered to be related to controlling force and modulating fatigue (Marsden et al. 2000). The strength of corticomuscular coupling is associated with motor behaviour, such as accuracy of task performance and force generated (Kilner et al. 2000, 2003). Several studies have reported significant CMC with the beta frequency range (15–35 Hz) during low to moderate sustained isometric contractions (Farmer et al. 1993; Halliday et al. 1998). Omlor et al. (2007) demonstrated that dynamic force output is accompanied by gamma-range (30–45 Hz) CMC, with a concomitant decrease in the beta-range CMC, thereby indicating that corticomuscular coupling is task-dependent.

## EVENT-RELATED POTENTIALS

ERPs represent the brain's neural response to specific sensory, motor and cognitive events. ERPs represent the outcome of signal averaging of EEG epochs time-locked to a particular stimulus or response event. The resulting waveform is then examined for changes in the amplitude and latency of different ERP *components*, typically defined as positive or negative peak activity (such as the P3 and N1 components), which refer to the first major negative and third major positive waves in the ERP waveform. Other ERP waves include slowly rising electrical activity such as the lateralized readiness potential. For an overview of these and other ERP components, see Luck (2005).

At the present time, ERPs hold a somewhat unique position in the tool shed of cognitive neuroscientists, because they provide the only neuroimaging technique that has high temporal resolution, of the order of milliseconds, compared to techniques such as PET and fMRI which are inherently sluggish (because they index cerebral haemodynamics). ERPs have a long history of use in cognitive neuroscience, where they have been examined in relation to cognitive functions such as attention, memory and language (Luck 2005). The use of ERP in cognitive ergonomics has focused on issues related to the assessment of human mental workload (see Chapter 18) and human error (Fedota and Parasuraman 2010).

ERP studies of physical work have typically focused on the motor-related cortical potential (MRCP). This ERP component was first discovered by Kornhuber and Deecke (1965). It is characterized by a slowly rising negative potential, called *Bereitschaftspotential* or readiness potential, which is followed by a sharp rising negative potential, known as negative slope. As the onset of MRCP occurs prior to the onset of the motor activity, MRCP is considered to indicate pre-motor activity, which involves specific brain regions that prepare for a desired motor behaviour. We discuss studies using the MRCP in more detail later in this chapter.

## NON-INVASIVE BRAIN MODULATION

All of the methods we have discussed thus far are based on sensing brain activity while a human operator is engaged in cognitive or physical work. As such, all such techniques are *correlational*, so that any association between neural activity and work-related performance must be interpreted appropriately, typically with the use of control conditions. Even with good controls, however, it may be difficult to establish *causal* links between brain activity and performance using these methods. Researchers have therefore turned to non-invasive stimulation techniques that modulate brain



activity, such as TMS and tDCS, in order to establish such causal associations. These techniques allow for temporary inhibition or activation of specific brain regions. The logic of their use is as follows. If a neuroimaging study finds that brain region X is activated (relative to control) when a person is engaged in function Y, then the *necessity* of X for Y can be shown if temporary inhibition of X (by TMS or tDCS) reduces the efficiency of Y, or if transient activation of X improves performance. This procedure has been used in several studies examining the causal role of different brain regions in various cognitive functions (Walsh and Pascual-Leone 2005).

TMS and tDCS can also be used to modulate brain activity so that the performance of a given cognitive or motor task is improved. Alternatively, these techniques can be applied not to enhance performance over baseline, but to reduce or eradicate a normally occurring performance limitation. For example, tDCS can be used to reduce the normal decrement over time that occurs in vigilance tasks (Nelson et al. 2014).

TMS systems include a magnetic coil that is positioned over the participant's scalp over a brain region of interest. An electrical current is passed through the coil for very brief periods of time, creating a changing magnetic field perpendicular to the head. The magnetic field passes through the skull and induces current flow in the underlying cortical issue, sufficient to alter neural firing (Walsh and Pascual-Leone 2005). The spatial resolution of TMS is relatively high, particularly when the participant's MRI scan is available to guide the TMS coil placement. The temporal resolution is also high, since the TMS pulses can be delivered with millisecond precision.

Whereas TMS uses changing magnetic pulses, tDCS uses small DC electric current with electrodes attached to the scalp. Typically, a constant small current of 1 or 2 mA is applied. A positive polarity (anode) is typically used to stimulate neuronal function and enhance performance. Conversely, a negative polarity (cathode) is used to inhibit neuronal activity. The mechanism by which tDCS influences brain function is not precisely known, but is thought to alter the electrochemical environment of neurons, so that they fire more readily to input from other neurons (Bikson et al. 2006). Compared to the TMS, tDCS has low spatial and temporal resolution, but has the advantage that it is portable and very inexpensive.

## CRITICAL ISSUES IN COGNITIVE NEUROERGONOMICS

Since its beginning (Parasuraman 2003), the neuroergonomic approach has been applied to a number of different issues in cognitive ergonomics. Space limitations do not allow us to provide a comprehensive coverage. Instead, we focus on a few key issues in cognitive work and on those areas where the most research and development work has been done. These include (1) mental workload, (2) vigilance and mental fatigue, (3) neuroadaptive systems and (4) training.

### ASSESSMENT OF MENTAL WORKLOAD

The assessment of human mental workload is one of the most widely studied topics in ergonomics (Wickens and McCarley 2008 and Chapter 18 of this book). Because unbalanced operator mental workload (either too high or too low) can be a restrictive element in the efficiency and safety of human-machine systems, workload must be assessed in the design of new systems or the evaluation of existing ones. Behavioural measures, such as accuracy and speed of response on secondary tasks, or subjective reports (such as the NASA Task Load Index (TLX)) have been widely used to assess mental workload. However, measures of brain function offer some unique advantages that can be exploited in mental workload assessment (Kramer and Parasuraman 2007). Among these is the ability to extract covert physiological measures continuously in complex system operations in which overt behavioural measures may be relatively sparse. Perhaps a more compelling rationale is that measures of brain function can be linked to emerging cognitive neuroscience knowledge on attention (Wickens and McCarley 2008), thereby allowing for the development of neuroergonomic theories that in turn can advance practical applications of research on mental workload.

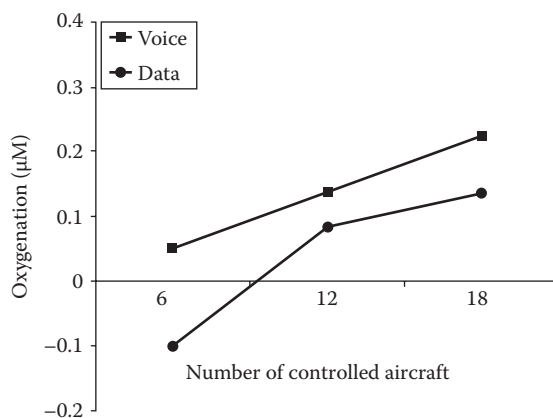


The dominant theory of human mental workload is *resource* theory (Wickens 1984, 2002). This theory postulates that except for highly overlearned *automatic* tasks, task performance is directly proportional to the application of attentional resources. In addition, the theory postulates that the pattern and amount of interference when two or more tasks are performed simultaneously (such as driving and talking on the cell phone) is related to the degree of overlap of multiple pools of resources.

Dual-task studies have provided abundant support for resource theory (Wickens and McCarley 2008). However, the theory has also been criticized for circularity (Navon 1984), one aspect of which can be linked to the lack of an independent measure of the resource construct. Neuroergonomic measures can negate this criticism if neural measures of mental resources can be identified.

fMRI findings showing that increased cerebral blood flow in regions of the PFC is associated with resource utilization have provided such validation (Parasuraman and Caggiano 2005). Other measures of cerebral haemodynamics, such as fNIRS and TCD, have also provided validation for the resource construct. In a recent study, Ayaz et al. (2012) tested experienced air traffic controllers on a complex task requiring them to keep aircraft in their sector free of conflicts. The number of aircraft under positive control was varied, and controllers used either traditional voice or digital text to communicate with pilots in the sector (voice and data communications, respectively). fNIRS was used to measure activation of the PFC as a measure of mental workload. Ayaz et al. (2012) found there was an increase in prefrontal cortical oxygenation as the number of aircraft in the sector increased from 6 to 18. As Figure 22.1 shows, there was a steady increase in blood oxygenation as aircraft load increased. At the same time, mental workload, as reflected in oxygenation, was lower for data than for voice communications. These findings suggest that fNIRS can provide a sensitive index of cognitive demand in a skilled group performing a realistic task that was highly representative of their work environment. Other recent studies have also used fNIRS to assess the effects of *think aloud* protocols (Pike et al. 2014) and to detect overload in complex human-computer systems (Maier et al. in press).

TCD has also been found to be sensitive to variations in mental workload. For example, Satterfield et al. (2012) asked participants to perform a simulated air defence task requiring control of multiple unmanned vehicles. The task load was varied by increasing or decreasing the number of enemy aircraft attempting to penetrate the defensive area. Consistent with the view that TCD-measured blood flow changes reflect attentional resource availability, blood velocity increased dynamically with higher task load and decreased with lower task load.



**FIGURE 22.1** Change in blood oxygenation in the prefrontal cortex as a function of the number of aircraft under positive control in the ATC's sector and with voice or data communications. (Adapted from Ayaz, H. et al., *NeuroImage*, 59, 36, 2012.)

Another aspect of Wickens' (1984) resource theory is that different types of resources are used in the performance of different tasks. Evidence for this aspect of the theory has been obtained primarily from dual-task studies in which different patterns of dual-task performance decrement are found, depending on the task pairings. Here too, neuroergonomic measures have provided converging evidence for the theory. For example, fMRI findings (Just et al. 2003) have supported the distinction among perceptual/cognitive, verbal/spatial and focal/ambient visual processing, which are components of Wickens' (1984) multiple resource model (updated in Wickens 2002).

Additional support for the multiple resource theory has come from ERP studies. One ERP component, the P300, has been examined in studies whose results have contributed to an understanding of the structure of mental workload. P300 is a positive ERP wave typically elicited by low-probability targets interspersed with more frequent non-targets. P300 latency increases with the difficulty of identifying targets but not with increases in the difficulty of response choice, suggesting that P300 provides a relatively pure measure of perceptual processing/categorization time, independent of response selection/execution stages (Kutas et al. 1977). P300 amplitude is also proportional to the amount of attentional resources allocated to the target (Johnson 1986). Thus, any diversion of resources away from target discrimination in a dual-task situation will lead to a reduction in P300 amplitude.

In a dual-task study, Isreal et al. (1980) showed that P300 amplitude decreased when a primary task, tone counting, was combined with a secondary task of visual tracking. However, increases in the difficulty of the tracking task did not lead to a further reduction in P300 amplitude. These findings showed that P300 reflects processing resources associated with perceptual processing and stimulus categorization, but not response-related processes, and provides strong validation of the distinction made between central and response-related resources in the multiple resource model workload.

ERP measures have also contributed to mental workload assessment in complex tasks. For example, Schultheis and Jamieson (2004) used secondary task P300 as a tool to assess the difficulty of a hypermedia educational system for individual learners, hoping to adapt difficulty on a person-by-person basis by assessing individual workload demands. They found that secondary task P300 amplitude was sensitive to the difficulty of text presented in a hypermedia system. They concluded that P300 amplitude and other measures, such as reading speed, may be combined to evaluate the relative ease of use of different hypermedia systems. For another example from the domain of driving assessment, Baldwin and Coyne (2005) found that P300 amplitude was sensitive to the increased difficulty of simulated driving in poor visibility due to fog, compared to driving in clear conditions. The value of this neuroergonomic measure was shown by the finding that performance-based and subjective indices were not affected by visibility conditions.

In addition to ERP measures such as P300 amplitude, spectral power in different EEG frequency bands has also been found to be sensitive to different aspects of mental workload. Specifically, frontal theta activity (4–7 Hz) increases while alpha power (8–12 Hz) decreases as the resource demands of a task increase (Gevins and Smith 2003). Thus, these EEG measures are well suited to the assessment of operator mental workload, not just in the laboratory but also in operational environments such as flight, air traffic control (ATC) and road and rail transportation (Brookhuis and De Waard 2002; Wilson 2001, 2002).

There are many factors, such as cost, ease of implementation and intrusiveness, that must be taken into consideration when selecting neuroergonomic techniques for mental workload assessment. Some of these factors (e.g. cost) may rule out the use of neuroergonomic methods in favour of simpler indexes such as subjective measures. Some individuals may also not wish to be *wired up* for physiological recording in work environments, so operator acceptance must also be carefully considered. However, with increasing miniaturization and development of *dry electrode*, wireless, wearable systems, some of these concerns are diminishing. At the same time, even if practical considerations rule out the use of physiological measurement, the neuroergonomic approach may nevertheless remain important for theory development, which in turn may lead to the development of more sensitive measures of mental workload (Kramer and Parasuraman 2007).

## VIGILANCE AND MENTAL FATIGUE

The evaluation of operator vigilance and mental fatigue in work environments is a topic closely related to workload assessment. The widespread implementation of automation in many work environments, including air and surface transportation and health care, while often leading to a reduction in operator workload, can also increase workload because of the resulting need for monitoring the automation (Parasuraman 1987). Furthermore, long work hours can further increase demands on operator vigilance, with the potential consequence of inducing mental fatigue.

The typical finding in studies of vigilance is that the detection rate of critical targets declines with time on task (Davies and Parasuraman 1982). This *vigilance decrement* was originally attributed to a reduction in physiological arousal due to the relatively monotonous nature of vigilance tasks (Frankmann and Adams 1962). However, more recent research indicates that the performance of vigilance tasks requires considerable cognitive work and induces stress reactions (Warm et al. 1996, 2008). There are two sources of evidence for these assertions, the first based on subjective measures and the second on neuroergonomic measures.

Performing a vigilance task is tedious and can induce boredom, particularly if the work period is prolonged. However, when observers' subjective perceptions are examined, a vigilance task appears to be a high workload assignment, with the ratings often being in the upper end of the NASA-TLX scale (Warm et al. 1996). Warm et al. (1996) conducted a series of studies showing that rather than being under-arousing, vigilance tasks are resource demanding and associated with high mental workload. Furthermore, they found that the vigilance decrement is accompanied by a linear increase in overall workload over time and that overall workload is closely tied to the level of task demand imposed by the vigilance task.

The second source of evidence associating vigilance performance with high workload and stress comes from neuroimaging studies. In recent years, studies using PET and fMRI have shown that changes in cerebral blood flow and glucose metabolism are involved in the performance of vigilance tasks (Parasuraman et al. 1998). These studies have also identified several brain regions that are active in such tasks, particularly the right prefrontal and parietal cortices, as well as subcortical nuclei such as the locus coeruleus. However, these studies have typically not linked these regions to performance changes in the vigilance task, particularly with respect to the vigilance decrement, which refers to the decline in detection efficiency or speed of response with time on task (for an exception, see Lim et al. 2010). Moreover, because the MRI testing environment is very restrictive, requiring observers to keep motionless and exposing them to loud noise, both of which affect vigilance behaviour, fMRI may not be ideal for linking changes in brain function with vigilance performance over a prolonged period of time. Two less restrictive measures, TCD and fNIRS, have accordingly been used more extensively in neuroergonomic studies of vigilance (Warm and Parasuraman 2007).

Warm and colleagues reported a series of studies of TCD and vigilance (for reviews, see Warm and Parasuraman 2007; Warm et al. 2008). A consistent finding is that the vigilance decrement is paralleled by a decline in blood flow velocity over time, relative to a baseline of activity just prior to beginning the vigilance session. The parallel decline in vigilance performance and in blood flow velocity is found for both visual and auditory tasks (Shaw et al. 2009). These findings have been interpreted using resource theory, which was described earlier. A critical control finding in support of the resource theory – as opposed to a generalized arousal or fatigue model – is that the blood flow change occurs only when observers actively engage with the vigilance task. When observers are asked to simply watch a display passively without having to detect targets for the same amount of time as in an active vigilance condition – a case of maximal under-arousal – blood flow velocity does not decline but remains stable over time.

How can the deleterious effects of loss of operator vigilance be countered? One method is to reduce work hours and increase rest breaks, but this may not be practical in all work settings. Furthermore, if the vigilance task is sufficiently demanding, vigilance decrements can occur in

as little as 5 min of work (Nuechterlein et al. 1983). Thus, other mitigating strategies need to be pursued. One method is to use cueing. Detection performance in vigilance tasks can be improved by providing observers with consistent and reliable cues to the imminent arrival of critical signals, with the extent of the decrement being reduced or eliminated (Wiener and Attwood 1968). With cueing, observers need to monitor a display only after having been prompted about the arrival of a signal and therefore can husband their information processing resources over time. In contrast, when no cues are provided, observers are never certain of when a critical signal might appear and consequently have to process information on their displays continuously across the watch, thereby consuming more of their resources over time than cued observers. Such cueing effects are particularly effective when signals are difficult to discriminate because of poor viewing conditions. For example, Parasuraman et al. (2009) showed that observers exhibited a vigilance decrement when required to discriminate critical signals representing particular combinations of hand movements with objects in video scenes (e.g. a hand picking up a gun in order to fire it vs. a similar movement in order to transport the gun), but only when the videos were visually degraded. Under such degraded viewing conditions, pre-cueing the critical signals reduced the vigilance decrement, even when the cues were not perfectly predictive (e.g. 70% reliable). In operational settings, cues could be generated from automated systems for detecting specific types of movements from analysis of video images from surveillance camera (e.g. Wang et al. 2003). Such systems are typically not 100% reliable, but as the results of Parasuraman et al. (2009) showed, even imperfect aiding can boost detection performance in tasks requiring detection of threats over a prolonged work period.

If the vigilance decrement stems from resource depletion due to the need to attend continuously to a display, then pre-cues should reduce the decline in cerebral blood flow velocity as measured by TCD, as confirmed in a study by Hitchcock et al. (2003). They used no pre-cues or pre-cues that were 100%, 80% or 40% reliable in pointing to an upcoming critical event in a simulated air traffic control task. Performance efficiency remained stable in the 100% reliable cueing condition but declined over time in the remaining conditions, so that by the end of the vigil, performance efficiency was clearly best in the 100% group, followed in order by the 80%, 40% and no-cue groups. Blood flow declined in the no-cue control condition, but there was a progressive reduction in the extent of the decline with progressively more reliable cues. There was no decline when the cues were perfectly reliable. This pattern of change in blood flow exactly matched that of performance.

In addition to cueing, non-invasive brain stimulation could also be used to mitigate vigilance decrement and mental fatigue. Nelson et al. (2014) applied 1 mA anodal tDCS to either the left or right PFC while participants performed the same vigilance task used by Hitchcock and colleagues. tDCS was applied either early or late during the course of the vigilance task. Compared to a control group that showed the normal vigilance decrement, the early stimulation group had a higher detection rate of critical signals. The late stimulation group initially exhibited a vigilance decrement, but this was reversed following application of tDCS. These initial findings are highly encouraging, but need to be followed up with additional research to examine the long-term effectiveness of tDCS as a method to alleviate vigilance problems at work.

One of the goals of neuroergonomics is to enhance understanding of aspects of human performance in complex systems with respect to the underlying brain mechanisms and to provide measurement tools to study these mechanisms (Parasuraman 2003). From this perspective, the use of TCD-based measures of cerebral blood flow to assess human mental workload and vigilance, and the effects of non-invasive brain stimulation, can be considered a success. The vigilance studies have revealed a close coupling between vigilance performance and blood flow, and they provide empirical support for the notion that blood flow may represent a metabolic index of information processing resource utilization.

To date there have been relatively few neuroergonomic investigations of the assessment and/or mitigation of vigilance problems in real settings. One recent example is the assessment of vigilance as it is affected by the length and quality of sleep in everyday life. Gartenberg and Parasuraman (2010) provided users with a mobile phone application based on the psychomotor

vigilance task (PVT). The PVT is widely used in studies to assess the effects of fatigue and sleep deprivation on human performance (Mallis et al. 2007). However, Gartenberg and Parasuraman (2010) wanted to develop a version of the task that could be used by individuals in their daily routine as a way to assess their sleep quality. The use of the mobile phone provided a unique opportunity for collecting sleep-related data under naturalistic conditions (see also Gartenberg et al. 2012). In addition to a version of the PVT, the device included an alarm clock to track sleep duration and time of sleep and a diary that displayed sleep statistics to the user. Three versions of the PVT that varied in length were administered: 1, 2.5 and 5 min. Reaction time measures on all three tasks were sensitive to sleep deprivation, showing that the application could be used in longitudinal or other neuroergonomic studies aimed at assessing the effects of sleep quality in naturalistic environments.

## NEUROADAPTIVE SYSTEMS

While the goal of ergonomic design is to avoid having workers exposed to extremes of workload and to loss of vigilance, this may not always be possible in certain work settings where unexpected events, equipment failures or other unanticipated factors lead to a transient increase in the task load imposed on the human operator, or long work hours impose demands on operator vigilance. Adaptive automation offers one approach to deal with these issues (Parasuraman 1987, 2000). In this approach, the allocation of functions to human and machine agents is flexible during system operations, with greater use of automation during high task load conditions or emergencies and less during normal operations, consistent with the approach of dynamic function allocation (Lintern, 2012). The adaptive automation concept has a long history (Parasuraman et al. 1992), but neuroergonomic methods for its implementation have been considered relatively recently (Inagaki 2003; Parasuraman 2003; Scerbo 2007).

Several methods to implement adaptive systems have been examined, including neuroergonomic measures to assess the operator's functional state (Byrne and Parasuraman 1996; Kramer and Parasuraman 2007; Parasuraman and Wilson 2008; Ting et al. 2010; Wilson and Russell 2007). Many studies have used EEG because of its ease of recording and (relative) unobtrusiveness (compared, say, to secondary tasks or subjective questionnaires). EEG also has the property of being a very high bandwidth measure, offering the possibility of sampling the human operator at up to about 30 Hz (Wilson and Russell 2003).

Workload adaptive systems need to assess operator state in real time, or near real time, so that task allocation or restructuring can be implemented in cases of overload or underload. A number of different statistical and machine learning techniques have been used for this purpose. These include discriminant analysis (Berka et al. 2004), artificial neural networks (ANNs) (Baldwin and Penaranda 2012; Wilson and Russell 2007), Bayesian networks (Wang et al. 2011) and fuzzy logic (Ting et al. 2010). These have been implemented in real time and typically provide accuracies of 70%–85%.

Wilson and Russell (2007) carried out a study that is illustrative of the neuroergonomic approach to workload adaptive systems. In their study, participants scanned radar images provided by unmanned air vehicles for targets of interest. The targets had to be identified before a specific weapon's release point was reached. Different task loads (target search difficulty) were imposed so that the participants experienced varying levels of workload. An ANN was trained to discriminate between low and high periods of operator mental workload, based on inputs of EEG and other physiological variables. On detecting high workload, the ANN triggered adaptive aiding by slowing the speed of the vehicle associated with the target, thus providing additional time to complete the search before the vehicle reached the weapon release point. Such adaptive aiding led to a 50% improvement in targeting performance compared to a non-aiding condition. The effect on performance of workload-based adaptation was also greater than when the same amount of aiding was presented but randomly distributed, without being based on neuroergonomic measures.



Implementing neuroergonomic adaptive systems in real settings poses significant challenges. A major issue concerns the detection and removal in real time of artefacts (e.g. eye or body movements) from neural or other physiological measures of the operator's state. Furthermore, while initial success has been achieved in using computational techniques to classify workload on the basis of EEG and other neuroergonomic measures, the reliability and stability of these methods within and across individuals needs to be more rigorously tested (Christensen et al. 2012; Wang et al. 2011). Finally, the operational community must be involved in the design of adaptive systems to ensure user acceptance and compliance.

## TRAINING

The final topic we discuss in our examination of cognitive neuroergonomics is training. The development of training methods and their evaluation have been significantly influenced by a large body of knowledge on skill acquisition, its maintenance (retention) and its transfer to related activities (Stammers and Patrick 1975). The neural mechanisms that support such learning and retention have been a topic of increasing interest in cognitive neuroscience. Much of the impetus for this work has come from emerging findings in basic neuroscience on brain *plasticity*, the ability to change due to new learning or experience (Huttenlocher 2002). While it has long been known that the human brain undergoes dynamic changes in structure and function from infancy to late adolescence, recent findings show that the *adult* human brain retains a capacity for plasticity, even well into old age (Greenwood and Parasuraman 2012).

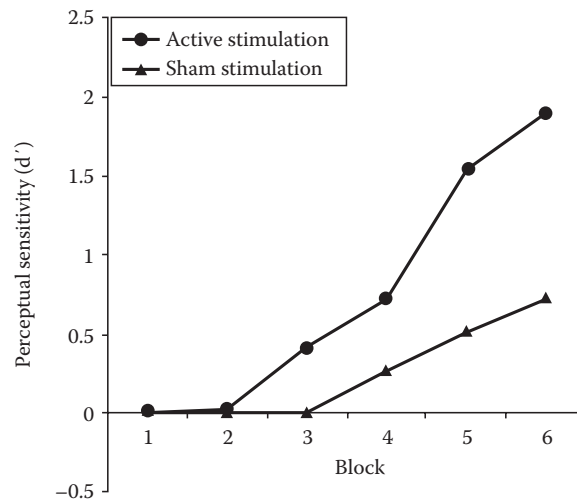
A particularly compelling example of brain plasticity comes from a study of people learning to juggle. Draganski et al. (2004) took structural MRI scans of participants before and after 3 months spent practising juggling a three-ball cascade. Compared to a control group that did not learn to juggle, the juggling group showed increases in the size of two brain regions important in motor control and coordination, the middle temporal area and the intraparietal sulcus. Thus, just like a muscle, the size of a brain area can increase with extended practice and training. The juggling group was scanned a third time after another 3 months, during which time they did not juggle. The brain regions that had expanded with practice had now reverted, although not completely, to their original size. These findings provide impressive evidence for the bi-directionality of brain plasticity.

Several recent studies have examined the effects of cognitive and motor training on white matter integrity as assessed by DTI. For example, Scholz et al. (2009) found improvements in FA following extended training in the same three-ball juggling task of Draganski et al. (2004). Similar increases in FA have been found after extended piano practice (Bengtsson et al. 2005) or playing golf (Jäncke et al. 2009). Similar benefits have been found in non-motor tasks, such as in DTI studies of extended practice on working memory tasks (Takeuchi et al. 2010). To date, there are relatively few DTI studies in neuroergonomics. The technique is well suited to examining training-related white matter changes associated with extended training in complex work tasks, and we expect that more such studies will be forthcoming. Training studies with DTI could also be conducted with high-fidelity simulations of natural environments.

Brain modulation techniques also have important implications for the neuroergonomics of training. TMS and tDCS have been shown to enhance performance in different perceptual, cognitive and motor tasks (Coffman et al. 2014; Utz et al. 2010). Most of these studies have examined modulatory effects on fairly simple tasks in a single session of performance. To address the potential of brain modulation as a training method, however, studies with more complex tasks and multiple sessions are needed to examine whether tDCS can accelerate learning.

Falcone et al. (2012) recently reported such a study. They had participants perform a complex military threat detection task requiring the detection of threats such as a concealed roadside bomb or a hidden sniper. The threats were very difficult to identify and participants initially performed at chance levels. Two groups were compared, one receiving active 2 mA tDCS and a control group receiving *sham* 0.1 mA stimulation, which was not sufficient to influence neuronal excitability.





**FIGURE 22.2** Changes in perceptual sensitivity ( $d'$ ) of threat detection over blocks with active (2 mA tDCS) or sham stimulation (0.1 mA). The first two blocks were baseline; stimulation (active or sham) was applied over the last four blocks. (Adapted from Falcone, B. et al., *PLoS One*, 7, e34993, 2012.)

To assess retention, participants were tested without tDCS immediately after the first session and again in a second session a day later. Anodal stimulation of right inferior frontal cortex with 2 mA increased perceptual sensitivity ( $d'$ ) on the task, compared to the 0.1 mA control, with no effect on response criterion  $\beta$ . As Figure 22.2 shows, participants in the 2 mA group had a much steeper learning curve compared to the control group. On completion of training, these participants had more than double the perceptual sensitivity of the control group. Furthermore, the performance enhancement was maintained for 24 h. These findings indicate that tDCS augments both skill acquisition and retention in a complex detection task and that the benefits are rooted in an improvement in sensitivity ( $d'$ ), rather than changes in response bias ( $\beta$ ). The results are encouraging with respect to the training effectiveness of tDCS, but additional work is needed to examine such issues as long-term retention, transfer and possible tolerance effects.

## CRITICAL ISSUES IN PHYSICAL NEUROERGONOMICS

From the beginning, neuroergonomics has been considered in relation to both cognitive (Parasuraman 2003) and physical work (Karwowski et al. 2003). Physical ergonomics focuses on human physical capabilities and limitations, pertaining to anthropometry, physiology and biomechanics of the human body, as they relate to physical work (Karwowski et al. 2003). Evaluation of physical work has traditionally been performed using the knowledge gained in the field of physical ergonomics. However, with the adoption of automation and human-computer design in work systems, it has become increasingly important to integrate physical, cognitive and affective capabilities and limitations in the design process. Physical neuroergonomics is an emerging field of study that focuses on the knowledge of human brain activities in relation to the control and design of physical tasks (Karwowski et al. 2003; Mehta and Parasuraman 2013).

Physical neuroergonomics emphasizes the importance of cognitive processing in the brain during motor activities, which is largely disregarded in traditional evaluation of physical work. For example, muscle fatigue is commonly defined as the inability to maintain a required force level after prolonged use of the muscle (Latash et al. 2003). Processes contributing to muscle fatigue can be divided into peripheral or central fatigue. Peripheral fatigue is associated with accumulation of metabolites and changes in the contractile elements, whereas central fatigue relates to changes

in the central nervous system that drives motor behaviour (Enoka and Duchateau 2008; Gandevia 2001; Taylor and Gandevia 2008). For example, factors such as lack of motivation, stress and mental fatigue have been shown to increase central fatigue (Boksem et al. 2006; Chaudhuri and Behan 2000). Unfortunately, traditional evaluations focus on biomechanical and physiological changes to quantify muscle fatigue and fail to consider the contributions of changes observed in the brain activity during fatigue manifestation. Concurrent assessment utilizing neuroergonomic and physical ergonomic principles may provide a better understanding of the underlying central and peripheral mechanisms of muscle fatigue. In this section, we focus on the applications of neuroergonomic techniques that have been employed to evaluate physical capabilities. These include evaluation of (1) physical work parameters, such as the level of force production and types of exertions; (2) muscle fatigue and (3) implications for health and performance.

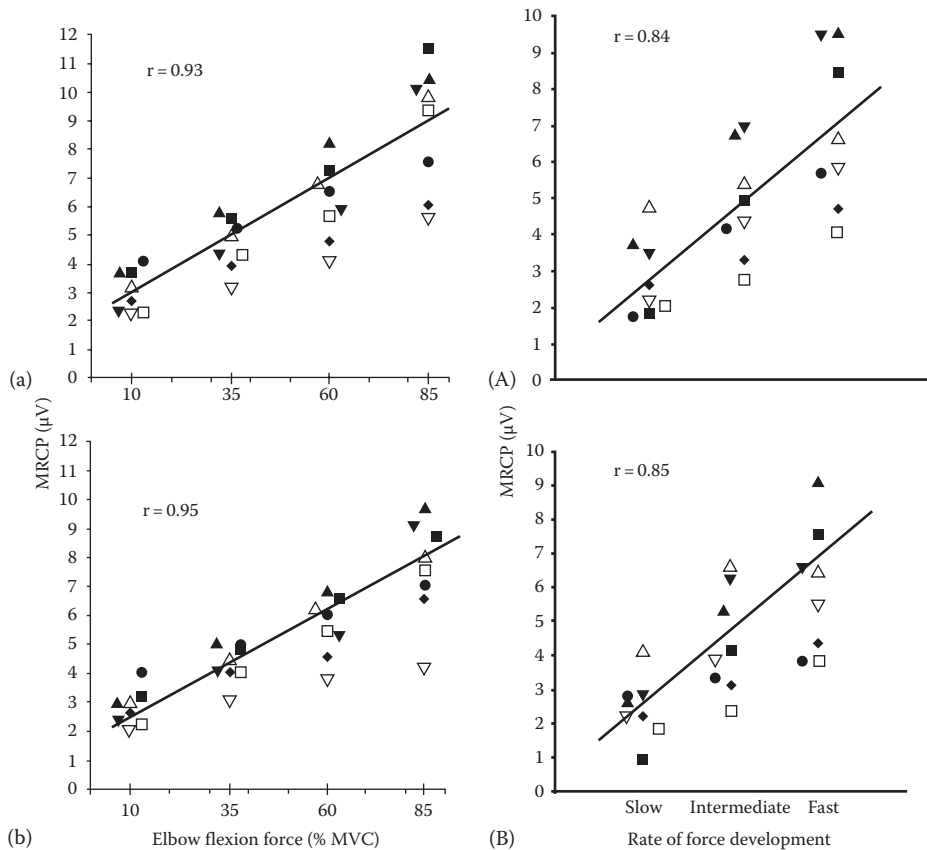
## **ASSESSMENT OF PHYSICAL WORK**

Assessment of physical work using neuroergonomic techniques is an emerging field of study that facilitates evaluation of human work, which may involve components of physical and cognitive information processing, and relates it to worker capacity, reflecting central and peripheral resources. Several characteristics of physical work can influence the type of measurement technique adopted. For example, employing the fMRI technique to understand cerebral oxygenation changes during dynamic tasks, such as repetitive lifting, is futile or even impossible to accomplish. Of the previously discussed neuroergonomic techniques, EEG and fNIRS offer many advantages that allow for the evaluation of a variety of physical tasks. EEG provides excellent temporal resolution, making it an invaluable tool to study motor behaviour at all stages (initiation, planning and execution) across different types of tasks. Moreover, effective artefact removal techniques are available that allow for its use in evaluating dynamic tasks. fNIRS provides good temporal resolution; it permits evaluation of movement-based motor activities, such as cycling, without the issues of artefact removal. Due to these advantages, the number of studies using EEG and fNIRS to investigate a variety of physical work (such as the level and type of physical exertion) is rapidly increasing. Work parameters, such as the levels of force production (low, moderate or high) and types of exertion (static or dynamic and flexion or extension), can influence human capabilities (both brain and body) differentially. The following section covers the evaluation of these work parameters using a physical neuroergonomic approach, focusing particularly on the application of EEG and fNIRS techniques.

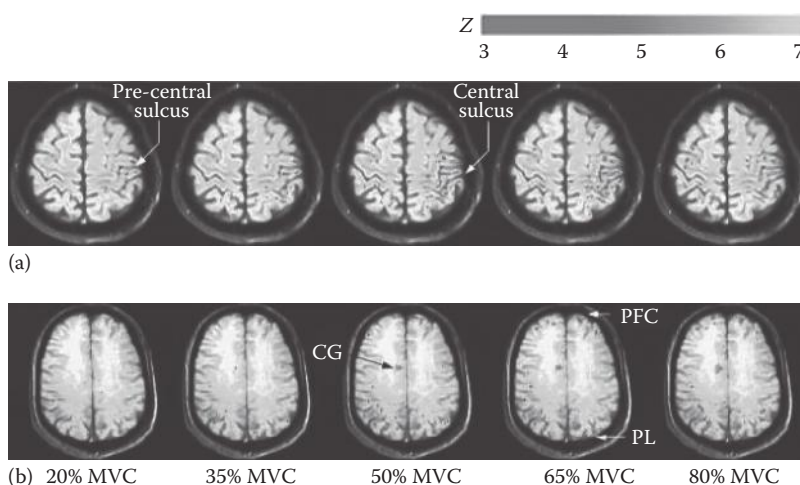
### **Force Production**

EEG-derived MRCP has been shown to be highly correlated with force production and rate of force production and associated muscle activity. Siemionow et al. (2000) demonstrated that MRCP obtained from the supplementary motor area and the contralateral sensorimotor cortex was correlated with muscle activity from the biceps brachii and joint force during isometric elbow-flexion contractions at four intensity levels (Figure 22.3a). They also reported that MRCP was highly correlated to the rate of force production at three different rates and associated muscle activity at a given elbow-flexion force level (Figure 22.3b). This study systematically studied the role of cortical motor commands (represented by the MRCP) on the control of voluntary muscle activation.

Given its high spatial resolution, fMRI has also revealed the contributions of the primary sensorimotor cortex in controlling motor behaviour and simultaneous activation of multiple cortical sites in planning and executing voluntary motor activities (König and Engel 1995). For example, Dai et al. (2001) investigated the relationship between fMRI-measured BOLD signal, force production and associated EMG. Participants performed five handgrip exertions, at 20%, 35%, 50%, 65% and 80% maximum voluntary contractions (MVCs) for ~30 s in each exertion level. Force and EMG data from the finger flexor and extensor muscles were also collected. The authors reported a direct relationship between handgrip force and fMRI signal and between muscle EMG and fMRI signal in several motor function-related cortical fields (Figure 22.4a and b).



**FIGURE 22.3** Relationship between MRCP and force (a) and rate of force production (b). (A) MRCP from the Supplementary Motor Area (SMA) location and (B) MRCP from the site of the sensorimotor cortex. Each symbol represents a subject ( $n = 8$ ). (After Siemionow, V. et al., *Exp. Brain Res.*, 133, 303, 2000. Reprinted with permission.)

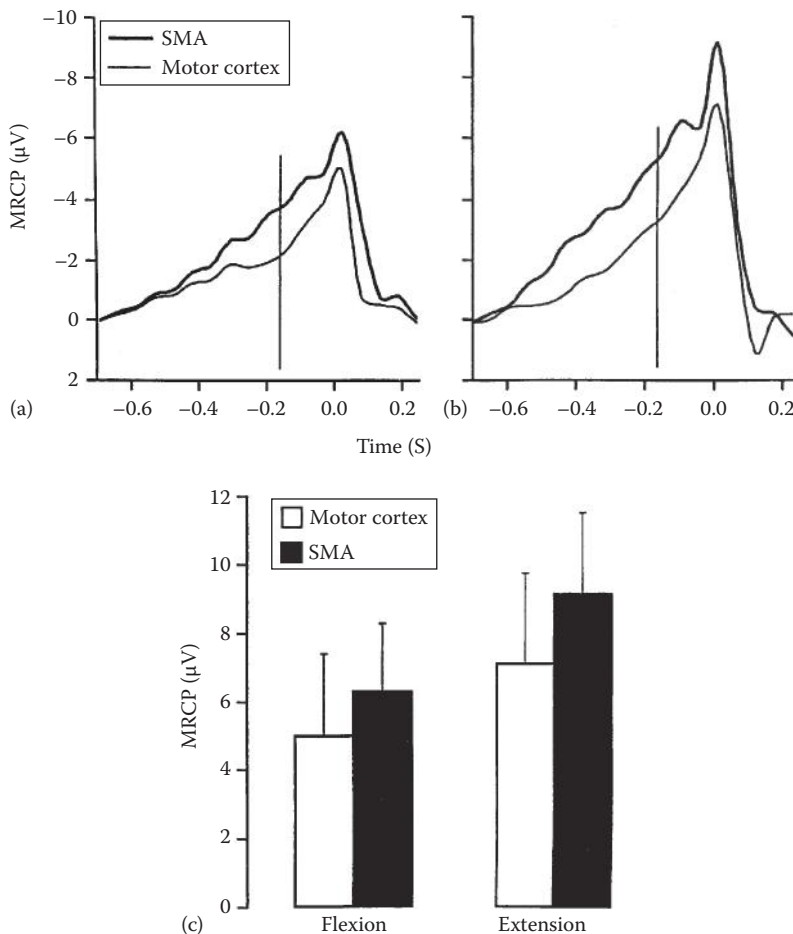


**FIGURE 22.4** Functional images (at the transverse plane) taken at the same brain locations at the five force levels. (a) The cortical level showing activation of the primary motor, sensory and pre-motor cortices. (b) Activities of the prefrontal (PFC), cingulate (CG) and parietal (PL) regions. (After Dai, T.H. et al., *Exp. Brain Res.*, 140, 290, 2001. Reprinted with permission.)

Stronger fMRI signals during higher force levels suggest that a greater number of cortical output neurons may participate in generating descending commands. Furthermore, the similarity in the relationship between motor output and fMRI signals from multiple cortical sites may indicate that a simultaneous activation across several brain regions may be necessary for precision force control.

### Muscle Flexion and Extension

Repetitive work tasks employ both flexion and extension movements across a given joint. Thus, it is important to understand the underlying central and peripheral mechanisms of movement control during such tasks. Yue et al. (2000) investigated the difference on the central nervous system control for thumb flexion and extension movements by comparing the amplitude of MRCP. EMG data from the flexor and extensor thumb muscles and EEG data from the supplementary motor area and contralateral motor cortex were collected from eight participants. The authors reported greater amplitude of MRCP during extension than flexion, despite a weaker extension EMG compared to the flexion EMG (Figure 22.5). Thus, using EEG-derived MRCP, it was concluded that stronger cortical activity occurred for planning and executing thumb extension movements



**FIGURE 22.5** Amplitude of motor-related cortical potential (MRCP) recorded during (a) the thumb flexion and (b) extension of one participant and (c) the group mean amplitude of the MRCP. Vertical lines in (a) and (b) indicate timing of beginning of movement, and time zero indicates timing of trigger. (After Yue, G.H. et al., *Brain Res.*, 856, 291, 2000. Reprinted with permission.)

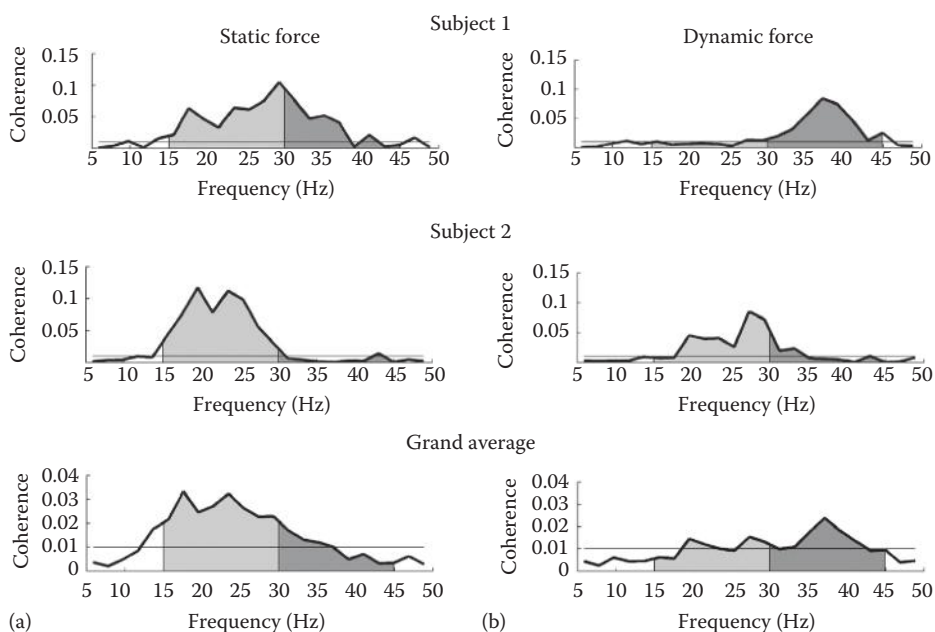
than for the flexion movement. The results suggested that extension movements might result from *separate* corticospinal projections to the motor neurons from those responsible for flexion movements.

### Static and Dynamic Exertions

Occupational tasks may contain components of static and dynamic exertions. Identifying neural mechanisms that contribute to these components is important in understanding motor performance. CMC between EEG and EMG can distinguish specific patterns of static and dynamic exertions. Omlor et al. (2007) compared EEG–EMG coherence in eight subjects during static and dynamic visuomotor tasks. Participants exerted a steady-state (static) or periodically modulated (dynamic) force output with their index finger to keep a visual cursor within a target zone. Steady-state force output was associated with CMC to the beta range. The dynamic conditions resulted in distinct coherence in the gamma range and significant reduction in the beta-range coherence (Figure 22.6). The authors concluded that during dynamic force, the corticospinal oscillation mode of the sensorimotor system shifts towards higher (gamma) frequencies to facilitate rapid integration of the visual and somatosensory information of the tracking task that is required to produce the appropriate motor command. Despite the involvement of the same muscles, coherence tends to shift to new frequencies when different tasks are performed (Marsden et al. 2000), reflecting its role in regulating fine and precise motor control.

### MUSCLE FATIGUE

Muscle fatigue is commonly defined as the inability to maintain a required force level after prolonged use of the muscle (Latash et al. 2003). Prolonged physical work, with inadequate rest breaks, can result in fatigue. While peripheral fatigue measures are widely used in physical ergonomics, neuroergonomic techniques can shed more light on the central mechanisms that contribute to



**FIGURE 22.6** Frequency–coherence plots for EEG–EMG coherence (a) during static force and (b) during dynamic force for subjects 1 and 2 and for all 8 subjects (grand average). The beta range (15–30 Hz) is marked in light grey and the gamma range (30–45 Hz) in dark grey. (After Omlor, W. et al., *NeuroImage*, 34, 1191, 2007. Reprinted with permission.)

fatigue development. As fatigue is a complex construct, it is not surprising that several EEG- and fNIRS-derived measures are available to evaluate fatigue and recovery. However, care should be taken to identify the type of task under study (force levels, static vs. dynamic, flexion vs. extension) when employing these measures.

EEG-derived MRCP has been previously employed to assess the nature (central or peripheral) of muscle fatigue. Johnston et al. (2001) compared MRCP associated with muscle fatigue with force and EMG activity during a grasping task. Six healthy participants maintained 70% of their maximum grip force for a total of 5 s in three blocks of 40 trials each. Peripheral fatigue was observed as the trials progressed, evident by decrease in force production and decrease in muscle activity, specifically in the last block. However, this block was associated with a significant increase in the activity of the *Bereitschaftspotential* (BP) component and the motor potential (MP) component of the MRCP. These increases in the early components of MRCP may reflect development of compensatory cortical strategies to accommodate for the inability to maintain the desired force levels due to peripheral fatigue. Ushiyama et al. (2011) found similar increases in CMC during sustained submaximal isometric contractions of the tibialis anterior muscle at 50% MVC. They attributed this increase in brain and muscle signal to additional strategy used by the central nervous system to accomplish the sustained contraction. In contrast, Yang et al. (2009) found that the EEG–EMG coherence decreased significantly from minimum fatigue to a stage of more severe fatigue during submaximal elbow flexion at 30% MVC. An explanation for these differences may lie in the type of task and muscles recruited. Activation of caudal and distal muscles has shown greater magnitude of EEG–EMG coherence than cranial and proximal muscles (Ushiyama et al. 2010). Moreover, task-related differences in the experimental protocols across these studies may further influence the reported findings. Thus, further investigation is warranted to clarify task- and muscle-specific changes in CMC during fatiguing contractions.

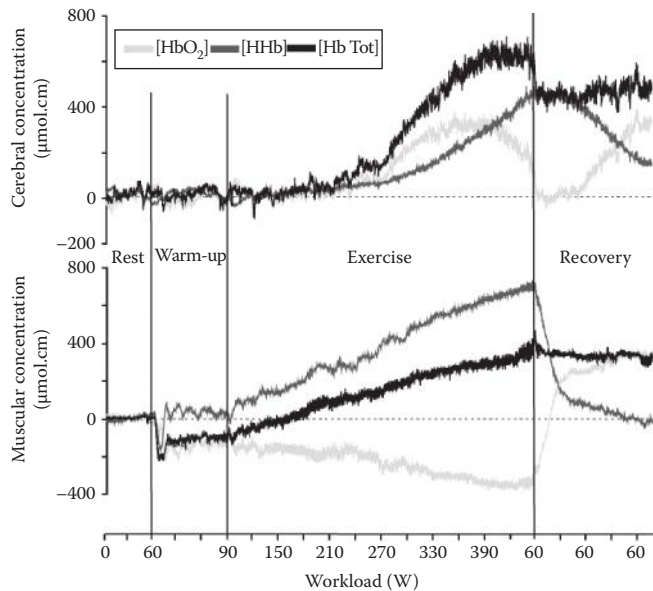
fNIRS-derived measures, such as cerebral oxygenation and individual components of haemoglobin content, have recently gained attention in monitoring central changes in oxygenation during exercise and physical work. Several studies have reported a significant decrease in prefrontal oxygenation, accompanied by muscular impairment, at exhaustion during submaximal and maximal fatiguing contractions (Bhambhani et al. 2007; González-Alonso et al. 2004; Nybo and Rasmussen 2007). Thomas and Stephane (2008) compared the changes in oxygenation of PFC and vastus lateralis muscle during incremental cycling exercise. Cerebral oxygenation increased in the early stages, but decreased markedly in the last stage until exhaustion, whereas muscle oxygenation decreased significantly throughout the exercise (Figure 22.7). The findings suggest that PFC activation contributes to the reduction in motor output at the cessation of exercise. However, further investigation is required to explore the causes of central fatigue using fNIRS-derived measures, as factors such as time delays and task dependencies may influence this relationship.

## IMPLICATIONS FOR HEALTH AND PERFORMANCE

Evaluation of physical work has traditionally been performed using the knowledge gained in the field of physical ergonomics. In order to understand how human physical capabilities are influenced by their type of work, it is important to assess brain functioning along with more peripheral measures. Physical neuroergonomics considers the contributions of cognitive processing in the brain during motor activities, which has led to a better understanding of underlying mechanisms of motor behaviour, particularly during fatigue.

Findings from studies that investigate cortical activation during various levels of force production, using EEG- and fMRI-derived measures, confirm that increase in force production is associated with a greater number of cortical output neurons and simultaneous activation of multiple brain regions during precision control (Dai et al. 2001; Siemionow et al. 2000). Tasks with high precision control require greater corticomuscular coupling, particularly in the beta range (Kristeva-Feige et al. 2002). Dynamic tasks, or tasks that require visual information, have been shown to increase





**FIGURE 22.7** Simultaneous time course changes in concentration of cerebral (top) and muscular (bottom) oxyhaemoglobin ([HbO<sub>2</sub>]), deoxyhaemoglobin ([HHb]) and total haemoglobin ([HbTot]) caused by a progressive maximal cycling exercise. Values are presented for a representative subject. (After Thomas, R. and Stephane, P., *Eur. J. Appl. Physiol.*, 102, 153, 2008. Reprinted with permission.)

corticomuscular coupling at higher frequencies, indicating the adaptive role of cortical oscillations in rapidly integrating visual (or new) information with the somatosensory information (Marsden et al. 2000; Omlor et al. 2007). These findings have important implications for task analysis and design, particularly for work tasks that require visual feedback or fine or precise control of body motions. Moreover, an inability to achieve such synchronization limits an operator's capacity to perform the task safely and efficiently (Karwowski et al. 2003).

Manual repetitive tasks (e.g. typing during computer work and assembly line tasks) are associated with increased incidence of musculoskeletal disorders (Bernard 1997). While traditional approaches focus on the impact of repetitive motion on muscular responses, EEG-derived MRCP findings demonstrated that extension and flexion movements result from separate corticospinal projections to the motor neurons (Yue et al. 2000). Extension movements elicited a greater brain response but a weaker muscular response than flexion movements. This particular finding may be important to our understanding of the aetiology of musculoskeletal disorders due to repetitive motion.

Overexertion, or fatigue, remains the primary cause of worker disability. Fatigue is a complex, multifaceted phenomenon; isolating the brain and the muscle for separate investigations can only complicate our understanding of the underlying mechanisms. Examining the role of brain functioning during fatigue development is critical to extend our knowledge on the aetiology and potential mechanisms of fatigue. The experiments discussed earlier suggest that fatigue progression is accompanied by new compensatory cortical strategies to accommodate for loss of force-generating capacity. Increase in EEG-derived measures, such as MRCP and CMC, during peripheral fatigue development, indicates the role of the brain in maintaining motor performance. In turn, a decrease in cortical activation, as observed with fNIRS-derived measures, may suggest the development of central fatigue, that is, the inability of the brain to produce descending commands during motor tasks. Investigations that focus on the interactions of both central and peripheral fatigue on brain functioning may extend our knowledge on human physical capabilities and limitations at work. As such, these studies can help facilitate the design process for work tasks or ergonomic interventions that focus on preventing work-related muscle fatigue and injury.

Current trends in the evaluation of work have become more multidisciplinary, borrowing principles from the field of neuroscience (as discussed previously). However, research is required to assess the effectiveness of these techniques on work-related tasks. While the cognitive ergonomics domain has witnessed a recent influx of neuroergonomic techniques in the evaluation of mental workload, vigilance and human error (Parasuraman 2011), there are fewer neuroergonomic studies in the evaluation of physical work. The nature of physical work, that is, movements or respiration artefacts, has limited the application of neuroergonomic techniques to specific tasks (e.g. stationary or controlled tasks). Nonetheless, the knowledge that can be gained by adopting a multidisciplinary approach (i.e. neuroscience and ergonomics techniques) can be used to determine the capabilities of workers and demands of their jobs. Such an approach is critical to improve our understanding of the underlying complex mechanisms of musculoskeletal injury development.

## **INTERACTIONS BETWEEN PHYSICAL AND COGNITIVE WORK**

We have considered neuroergonomic studies of physical and cognitive work separately in this chapter. This was done largely out of convenience and because there are fewer studies on the latter. However, there are many interactions between physical and cognitive work, so that ideally they should be considered together when examining brain and behaviour at work (Marras and Hancock 2014). High cognitive demands can influence physical work; and physical activity can influence cognitive processing. As an example of the former, consider nurses in their daily work activities: they are often exposed to high physical and cognitive demands, which may comprise lifting during patient handling to decision making during drug administration activities. An example of the latter is the impairment in memory and related cognitive functions that follows long-duration physical work and dehydration (Tomprowski 2003). We briefly discuss cognitive–physical work interactions, focusing on neuroergonomic-based studies that have explored the effects of (1) physical work and fatigue on cognitive function and (2) cognitive work on physical capacity.

### **EFFECTS OF PHYSICAL WORK ON COGNITIVE FUNCTION**

Kamijo et al. (2004) investigated the influence of exercise intensity on cognitive function using the P300 ERP component. Participants performed a go/no-go reaction time task in a control condition and again after high-, medium- and low-intensity pedalling exercises. The P300 amplitude increased after medium-intensity pedalling exercise, decreased after high-intensity pedalling exercise and demonstrated no change after low-intensity pedalling exercise. The results suggested that exercise influenced the amount of attentional resources devoted to a given task and that the changes in P300 amplitude followed an inverted U-shaped behaviour of differences in exercise intensity. A recent study by Yanagisawa et al. (2010) employed fNIRS to monitor cortical activation related to changes in the Stroop interference test during moderate exercise. Consistent with the results discussed earlier for moderate-level exercises, the study reported enhanced cognitive performance with concomitant increase in the left dorsolateral PFC. The ERP- and fNIRS-based findings discussed here support existing performance-based studies, such as those reviewed in Tomporowski (2003), where differences in physical exercise influence cognitive function.

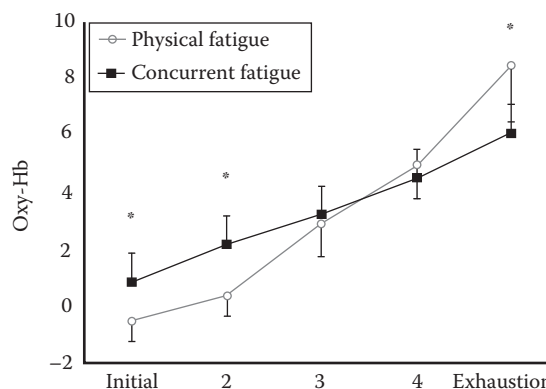
### **EFFECTS OF COGNITIVE WORK ON PHYSICAL CAPACITY**

Additional cognitive demand during physical work has been shown to adversely affect biomechanical responses during upper extremity and low back exertions (Marras et al. 2000; Mehta and Agnew 2011) and gait control while walking (Beauchet et al. 2003). A few studies have also reported adverse effects of cognitive demands on muscle capacity, specifically endurance, fatigue and recovery (Mehta and Agnew 2012; Yoon et al. 2009). Cognitive demand–related changes in muscle fatigue have been attributed, at least in part, to recruitment of the same motor units during

separate physical and mental work (Lundberg et al. 2002), decreased motor coordination (Mehta and Agnew 2011; Mehta et al. 2012), increased neuromotor noise (Bloemsaat et al. 2005; Gemmert and Galen 1997) and/or interference at the dorsolateral PFC that is involved in cognitive processing and isometric motor contractions (Dettmers et al. 1996; Rowe et al. 2000). Thus, in order to understand how human capabilities/limitations are influenced by physical and cognitive demands simultaneously, it is critical to assess neuromuscular alterations both at the cortical (central) and muscular (peripheral) levels.

Along with traditional physiological indicators (i.e. EMG), neuroergonomic techniques, such as EEG and fNIRS measurements to evaluate cortical changes, can provide information concerning the changes in the motor cortex under concurrent physical and cognitive stress. For example, Kristeva-Feige et al. (2002) employed EEG-derived coherence measures to evaluate corticomuscular coupling during motor tasks at varying precision levels in the absence and presence of a cognitive stressor. The study reported greater corticomuscular coupling at higher precision levels, particularly in the beta range (15–30 Hz). While additional cognitive demand did not change peripheral outcomes (i.e. EMG and force output) when compared to the control condition, a significant decrease in EEG–EMG coherence (i.e. beta-range synchronization) was observed. This decrease was attributed to the changes in the efficiency of the corticomuscular network when attention was divided between the motor and the cognitive tasks.

Recent years have seen the increased use of fNIRS to quantify both muscular (Mehta and Agnew 2012, 2013; Mehta et al. 2012) and cerebral tissue oxygenation (Ayaz et al. 2012; Shortz et al. 2012) during physical and cognitive work. However, there is limited evidence on the utility of fNIRS measurements during the evaluation of work that is characterized by *both* physical and cognitive demands. A recent study by Mehta and Parasuraman (2014) investigated the effects of combined physical and cognitive demands on PFC activity using near-infrared spectroscopy. Twelve healthy participants performed three fatiguing conditions: (1) handgrip at 30% MVC (physical fatigue), (2) mental arithmetic (mental fatigue) and (3) handgrip at 30% MVC + mental arithmetic (concurrent fatigue). Oxygenated haemoglobin (Oxy-Hb) was averaged across 10 s windows, at five stages of the endurance phase (initial phase, 25th% endurance time [ET], 50th% ET, 75th% ET and at exhaustion). Figure 22.8 illustrates the dynamic changes in Oxy-Hb across the physical and concurrent fatigue conditions. In general, Oxy-Hb increased until exhaustion during the physical and concurrent fatigue conditions; however, Oxy-Hb levels during the concurrent fatigue condition were higher in the initial stages but lower at exhaustion when compared to the physical fatigue condition despite comparable muscular output between the two conditions. Such reversal in relative magnitude of PFC activation between the two conditions highlights the influence of mental fatigue on central fatigue development.



**FIGURE 22.8** Mean oxygenated haemoglobin ( $n = 12$ ) during physical and concurrent fatiguing conditions at various phases of the ET. Error bars denote standard error. (Adapted from Mehta, R.K. and Parasuraman, R., *Hum. Factors*, 2014.)

Decreases in PFC have been associated with central fatigue during motor activities (Bhambhani et al. 2007); thus, the decrease observed in the study by Mehta and Parasuraman (in press) may be attributed to the synergistic demands of the handgrip and mental arithmetic tasks that may suggest interference at the PFC during fatigue manifestation. The results also indicated a decrease in mental arithmetic performance during the concurrent fatigue condition. Thus, it may be possible that decrease in cerebral oxygenation due to physical fatigue may have led to poor cognitive performance. However, a recent study by Ando et al. (2011) reported enhanced cognitive performance during moderate exercise, independent of cerebral oxygenation. Methodological differences may drive the discrepancies in the results obtained in these studies, as the latter study did not investigate muscle fatigue. Nonetheless, both studies highlight the potential of fNIRS-derived measures to evaluate brain functions during concurrent physical and cognitive work.

## CONCLUSIONS AND FUTURE DIRECTIONS

The integrated view of physical and cognitive work we have presented in this chapter is consistent with an emerging consensus in the philosophy of mind and body and its application to ergonomics (see also Marras and Hancock 2014). Research and practice in E/HF was dominated by behaviourism during its early history. There was little or no mention of the mind until the advent of cognitive psychology in the 1960s. At that time, however, little attention was paid to neuroscience or to the results of studies concerning brain mechanisms underlying human functions. The relative neglect of human brain function was consistent with a functionalist approach to the philosophy of mind (Dennett 1991). Such an approach implies that the characteristics of neural structure and functioning are largely irrelevant to the development of theories of mental functioning. However, the significance of findings from neuroscience has forced a retreat from this position. Furthermore, theories of *embodied* cognition (Clark 1998) propose that mind cannot be considered in isolation but in the context of physical action (as directed by our bodies under the control of the brain) and interaction with the artefacts and technologies of the everyday world. As such, the intent of this chapter was to highlight recent developments and adoption of neuroergonomic methods and applications, by integrating physical, cognitive and affective capabilities and their interactions, to evaluate human work.

Technical developments in neuroscience are providing ergonomics with a new set of tools to examine issues related to physical and cognitive work. Some of the techniques are cumbersome and expensive, such as the MRI, fMRI and DTI neuroimaging methods. These techniques will typically not be useful for applied ergonomic evaluations, where more portable methods, such as EEG, fNIRS and TCD, are more likely to be used. Nevertheless, neuroimaging methods may still have value in evaluation of different training methods. Furthermore, basic neuroimaging research can also inform ergonomic theories in different aspects of both physical and cognitive work. We discussed the application of neuroergonomic methods to several critical issues in the evaluation of physical and cognitive work. These included mental workload, vigilance and mental fatigue, neuroadaptive systems, training, physical exertions (i.e. force production, flexion or extension, static or dynamic) and muscle fatigue. As neuroergonomic methods, research and applications mature, we expect this list to cover many more issues that are of concern to ergonomists.

## REFERENCES

- Aaslid, R. 1986. Transcranial Doppler examination techniques. In *Transcranial Doppler Sonography*, ed. R. Aaslid, pp. 39–59. New York: Springer-Verlag.
- Ainslie, P.N., Barach, A., Murrell, C., Hamlin, M., Helleman, J. and Ogoh, S. 2007. Alterations in cerebral autoregulation and cerebral blood flow velocity during acute hypoxia: Rest and exercise. *American Journal of Physiology-Heart and Circulatory Physiology* 292: H976–H983.
- Ando, S., Kokubu, M., Yamada, Y. and Kimura, M. 2011. Does cerebral oxygenation affect cognitive function during exercise? *European Journal of Applied Physiology* 111(9): 1973–1982.

- Ayaz, H., Shewokis, P.A., Bunce, S. et al. 2012. Optical brain monitoring for operator training and mental workload assessment. *NeuroImage* 59: 36–47.
- Bailey, S.P., Hall, E.E., Folger, S.E. and Miller, P.C. 2008. Changes in EEG during graded exercise on a recumbent cycle ergometer. *Journal of Sports Science and Medicine* 7: 505–511.
- Baldwin, C.L. and Coyne, J.T. 2005. Dissociable aspects of mental workload: Examinations of the P300 ERP component and performance assessments. *Psychologia* 48(2): 102–119.
- Baldwin, C.L. and Penaranda, B. 2012. Adaptive training using an artificial neural network and EEG metrics for within-and cross-task workload classification. *NeuroImage* 59: 48–56.
- Beauchet, O., Kressig, R.W., Najafi, B. et al. 2003. Age related decline of gait control under a dual task condition. *Journal of the American Geriatrics Society* 51(8): 1187–1188.
- Bengtsson, S.L., Nagy, Z., Skare, S. et al. 2005. Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience* 8(9): 1148–1150.
- Berka, C., Levendowski, D.J., Cvetinovic, M.M. et al. 2004. Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction* 17(2): 151–170.
- Bernard, B.P. 1997. *Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-Related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back*. Cincinnati, OH: U.S. Department of Health and Human Services CDC (NIOSH) Publication No. 97-141.
- Bhambhani, Y., Malik, R. and Mookerjee, S. 2007. Cerebral oxygenation declines at exercise intensities above the respiratory compensation threshold. *Respiratory Physiology & Neurobiology* 156(2): 196–202.
- Bikson, M., Radman, T. and Datta, A. 2006. Rational modulation of neuronal processing with applied electric fields. *Proceedings of the IEEE Engineering in Medicine and Biology Society* 1: 1616–1619.
- Bloemsaat, J.G., Meulenbroek, R.G.J. and Galen, G.P.V. 2005. Differential effects of mental load on proximal and distal arm muscle activity. *Experimental Brain Research* 167: 622–634.
- Boksem, M.A.S., Meijman, T.F. and Lorist, M.M. 2006. Mental fatigue, motivation and action monitoring. *Biological Psychology* 72(2): 123–132.
- Brookhuis, K.A. and De Waard, D. 2002. On the assessment of (mental) workload and other subjective qualifications. *Ergonomics* 45(14): 1026–1030.
- Byrne, E.A. and Parasuraman, R. 1996. Psychophysiology and adaptive automation. *Biological Psychology* 42(3): 249–268.
- Calhoun, V. and Pearlson, G. 2012. A selective review of simulated driving studies: Combining naturalistic and hybrid paradigms, analysis approaches, and future directions. *NeuroImage* 59: 25–35.
- Callan, D., Gamez, M., Cassel, D.B., Terzibas, C., Callan, A., Kawato, M. and Sato, M. 2012. Dynamic visuomotor transformation involved with remote flying of a plane utilizes the ‘mirror neuron’ system. *PLoS One* 7(4): e33873. doi:10.1371/journal.pone.0033873.
- Chaudhuri, A. and Behan, P.O. 2000. Fatigue and basal ganglia. *Journal of the Neurological Sciences* 179(1–2): 34–42.
- Christensen, J.C., Estepp, J.R., Wilson, G.F. and Russell, C.A. 2012. The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage* 59: 57–63.
- Clark, A. 1998. *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.
- Coffman, B.A., Clark, V.P. and Parasuraman, R. 2014. Battery powered thought: Enhancement of attention, learning, and memory in healthy adults using transcranial direct current stimulation. *NeuroImage* 85: 895–908. dx.doi.org/10.1016/j.neuroimage.2013.07.083.
- Dai, T.H., Liu, J.Z., Sahgal, V., Brown, R.W. and Yue, G.H. 2001. Relationship between muscle output and functional MRI-measured brain activation. *Experimental Brain Research* 140(3): 290–300.
- Davies, D.R. and Parasuraman, R. 1982. *The Psychology of Vigilance*. London, U.K.: Academic Press.
- Dennett, D. 1991. *Consciousness Explained*. Cambridge, MA: MIT Press.
- Dettmers, C., Lemon, R., Stephen, K., Fink, G. and Frackowiak, R. 1996. Cerebral activation during the exertion of sustained static force in man. *Neuroreport* 7(13): 2103.
- Dickerman, R., McConathy, W., Smith, G., East, J. and Rudder, L. 2000. Middle cerebral artery blood flow velocity in elite power athletes during maximal weight-lifting. *Neurological Research* 22(4): 337.
- Draganski, B., Gaser, C., Busch, V. et al. 2004. Neuroplasticity: Changes in grey matter induced by training. *Nature* 427(6972): 311–312.
- Duschek, S. and Schandry, R. 2003. Functional transcranial Doppler sonography as a tool in psychophysiological research. *Psychophysiology* 40(3): 436–454.
- Enoka, R.M. and Duchateau, J. 2008. Muscle fatigue: What, why and how it influences muscle function. *The Journal of Physiology* 586(1): 11–23.



- Falcone, B., Coffman, B.A., Clark, V.P. and Parasuraman, R. 2012. Transcranial direct current stimulation augments perceptual sensitivity and 24-hour retention in a complex threat detection task. *PLoS One* 7(4): e34993.
- Farmer, S., Bremner, F., Halliday, D., Rosenberg, J. and Stephens, J. 1993. The frequency content of common synaptic inputs to motoneurons studied during voluntary isometric contraction in man. *The Journal of Physiology* 470(1): 127–155.
- Fedota, J.R. and Parasuraman, R. 2010. Neuroergonomics and human error. *Theoretical Issues in Ergonomics Science* 11(5): 402–421.
- Frankmann, J.P. and Adams, J. 1962. Theories of vigilance. *Psychological Bulletin: Psychological Bulletin* 59(4): 257.
- Gandevia, S.C. 2001. Spinal and supraspinal factors in human muscle fatigue. *Physiological Reviews* 81(4): 1725–1789.
- Gartenberg, D. and Parasuraman, R. 2010. Understanding brain arousal and sleep quality using a neuroergonomic smart phone application. In *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations*, eds. T. Marek, W. Karwowski and V. Rice, pp. 210–220. Boca Raton, FL: CRC Press.
- Gartenberg, D., Thornton, R., Masood, M., Pfannenstiel, D., Taylor, D. and Parasuraman, R. 2013. Collecting health-related data on the smart phone: mental models, cost of collection, and perceived benefit of feedback. *Personal and Ubiquitous Computing* 17(3): 561–570.
- Gartenberg, D., Thornton, R., Masood, M. et al. 2012. Collecting health-related data on the smart phone: Mental models, cost of collection, and perceived benefit of feedback. *Personal and Ubiquitous Computing* 1–10. doi: 10.1007/s00779-012-0508-3.
- Gemmert, A. and Galen, G. 1997. Stress, neuromotor noise, and human performance: A theoretical perspective. *Journal of Experimental Psychology Human Perception and Performance* 23: 1299–1313.
- Gevens, A. and Smith, M.E. 2003. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science* 4(1–2): 113–131.
- Gevens, A. and Smith, M.E. 2007. Electroencephalography (EEG) in neuroergonomics. In *Neuroergonomics: The Brain at Work*, eds. R. Parasuraman and M. Rizzo, pp. 15–31. New York: Oxford.
- González-Alonso, J., Dalsgaard, M.K., Osada, T. et al. 2004. Brain and central haemodynamics and oxygenation during maximal exercise in humans. *The Journal of Physiology* 557(1): 331–342.
- Greenwood, P.M. and Parasuraman, R. 2012. *Nurturing the Older Brain and Mind*. Cambridge, MA: MIT Press.
- Halliday, D.M., Conway, B.A., Farmer, S.F. and Rosenberg, J.R. 1998. Using electroencephalography to study functional coupling between cortical activity and electromyograms during voluntary contractions in humans. *Neuroscience Letters* 241(1): 5–8.
- Halliday, D.M., Rosenberg, J., Amjad, A. et al. 1995. A framework for the analysis of mixed time series/point process data-theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Progress in Biophysics in Molecular Biology* 64(2–3): 237–278.
- Hellström, G. and Wahlgren, N.G. 1993. Physical exercise increases middle cerebral artery blood flow velocity. *Neurosurgical Review* 16(2): 151–156.
- Hitchcock, E.M., Warm, J.S., Matthews, G. et al. 2003. Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science* 4(1–2): 89–112.
- Huttenlocher, P.R. 2002. *Neural Plasticity: The Effects of Environment on the Development of the Cerebral Cortex*. Cambridge, MA: Harvard University Press.
- Ide, K. and Secher, N.H. 2000. Cerebral blood flow and metabolism during exercise. *Progress in Neurobiology* 61(4): 397–414.
- Inagaki, T. 2003. Adaptive automation: Sharing and trading of control. In *Handbook of Cognitive Task Design*, ed. E. Hollnagel, pp. 46–89. Mahwah, NJ: Lawrence Erlbaum.
- Isreal, J.B., Chesney, G.L., Wickens, C.D. and Donchin, E. 1980. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology* 17(3): 259–273.
- Jäncke, L., Koeneke, S., Hoppe, A., Rominger, C. and Hänggi, J. 2009. The architecture of the golfer's brain. *PLoS One* 4(3): e4785.
- Johnson, R. 1986. A triarchic model of P300 amplitude. *Psychophysiology* 23(4): 367–384.
- Johnston, J., Rearick, M. and Slobounov, S. 2001. Movement-related cortical potentials associated with progressive muscle fatigue in a grasping task. *Clinical Neurophysiology* 112(1): 68–77.
- Jung, T.P., Makeig, S., Humphries, C. et al. 2000. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37(02): 163–178.



- Just, M.A., Carpenter, P.A. and Miyake, A. 2003. Neuroindices of cognitive workload: Neuroimaging, pupilometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science* 4(1–2): 56–88.
- Kamijo, K., Nishihira, Y., Hatta, A. et al. 2004. Differential influences of exercise intensity on information processing in the central nervous system. *European Journal of Applied Physiology* 92(3): 305–311.
- Karwowski, W., Siemionow, W. and Gielo-Perczak, K. 2003. Physical neuroergonomics: The human brain in control of physical work activities. *Theoretical Issues in Ergonomics Science* 4(1–2): 175–199.
- Kilner, J.M., Baker, S.N., Salenius, S., Hari, R. and Lemon, R.N. 2000. Human cortical muscle coherence is directly related to specific motor parameters. *The Journal of Neuroscience* 20(23): 8838–8845.
- Kilner, J.M., Salenius, S., Baker, S. et al. 2003. Task-dependent modulations of cortical oscillatory activity in human subjects during a bimanual precision grip task. *NeuroImage* 18(1): 67–73.
- König, P. and Engel, A.K. 1995. Correlated firing in sensory-motor systems. *Current Opinion in Neurobiology* 5(4): 511–519.
- Kornhuber, H.H. and Deecke, L. 1965. Changes in the brain potential in voluntary movements and passive movements in man: Readiness potential and reafferent potentials. *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere* 284: 1.
- Kramer, A. and Parasuraman, R. 2007. Neuroergonomics – Application of neuroscience to human factors. In *Handbook of Psychophysiology*, eds. J. Caccioppo, L. Tassinary and G. Berntson, 2nd edn. New York: Cambridge University Press.
- Krause, G., Ullsperger, P., Beyer, L. and Gille, H. 1983. Changes in EEG power density spectrum during static muscle work. *European Journal of Applied Physiology and Occupational Physiology* 51(1): 61–66.
- Kristeva-Feige, R., Fritsch, C., Timmer, J. and Lücking, C.H. 2002. Effects of attention and precision of exerted force on beta range EEG-EMG synchronization during a maintained motor contraction task. *Clinical Neurophysiology* 113(1): 124–131.
- Kubitz, K. and Mott, A. 1996. EEG power spectral densities during and after cycle ergometer exercise. *Research Quarterly for Exercise and Sport* 67(1): 91.
- Kutas, M., McCarthy, G. and Donchin, E. 1977. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science* 197(4305): 792–795.
- Lataash, M.L., Danion, F. and Bonnard, M. 2003. Effects of transcranial magnetic stimulation on muscle activation patterns and joint kinematics within a two-joint motor synergy. *Brain Research* 961(2): 229–242.
- Lim, J., Wu, W., Wang, J. et al. 2010. Imaging brain fatigue from sustained mental workload: An ASL perfusion study of the time-on-task effect. *NeuroImage* 49(4): 3426–3435.
- Lintern, G. 2012. Work-focused analysis and design. *Cognition, Technology & Work*, 14(1): 71–81.
- Luck, S.J. 2005. *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Lundberg, U., Forsman, M., Zachau, G. et al. 2002. Effects of experimentally induced mental and physical stress on motor unit recruitment in the trapezius muscle. *Work & Stress* 16(2): 166–178.
- Maguire, E.A. 2007. Spatial navigation. In *Neuroergonomics: The Brain at Work*, eds. R. Parasuraman and M. Rizzo, pp. 131–145. New York: Oxford University Press.
- Maikala, R.V. 2010. Modified Beer's law-Historical perspectives and relevance in near-infrared monitoring of optical properties of human tissue. *International Journal of Industrial Ergonomics* 40(2): 125–134.
- Maier, H.A., Pike, M., Wilson, M.L. and Sharples, S. 2014. Continuous Detection of Workload Overload: An fNIRS Approach. In *Contemporary Ergonomics and Human Factors 2014: Proceedings of the international conference on Ergonomics & Human Factors 2014*, Southampton, UK, 7–10 April 2014 (p. 450). CRC Press.
- Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. and Poizner, H. 2009. Linking brain, mind and behavior. *International Journal of Psychophysiology* 73: 95–100. doi: 10.1016/j.ijpsycho.2008.11.008.
- Mallis, M.M., Banks, S. and Dinges, D.F. 2007. Sleep and circadian control of neurobehavioral functions. In *Neuroergonomics: The Brain at Work*, eds. R. Parasuraman and M. Rizzo, Vol. 195177614, p. 207. New York: Oxford University Press.
- Marras, W.A. and Hancock, P.A. 2014. Putting mind and body back together: A human-systems approach to the integration of the physical and cognitive dimensions of task design and operations. *Applied Ergonomics* 45: 55–60.
- Marras, W.S., Davis, K.G., Heaney, C.A., Maronitis, A.B. and Allread, W.G. 2000. The influence of psychosocial stress, gender, and personality on mechanical loading of the lumbar spine. *Spine* 25(23): 3045–3054.
- Marsden, J., Werhahn, K., Ashby, P. et al. 2000. Organization of cortical activities related to movement in humans. *The Journal of Neuroscience* 20(6): 2307–2314.

- Mehta, R. and Parasuraman, R. 2013. Neuroergonomics: A review of applications to physical and cognitive work. *Frontiers in Human Neuroscience* 7: 889. doi: 10.3389/fnhum.2013.00889.
- Mehta, R.K. and Agnew, M.J. 2011. Effects of concurrent physical and mental demands for a short duration static task. *International Journal of Industrial Ergonomics* 41(5): 488–493.
- Mehta, R.K. and Agnew, M.J. 2012. Influence of mental workload on muscle endurance, fatigue, and recovery during intermittent static work. *European Journal of Applied Physiology* 112(8): 2891–2902.
- Mehta, R.K. and Agnew, M.J. 2013. Exertion-dependent effects of physical and mental workload on physiological outcomes and task performance. *IIE Transactions on Occupational Ergonomics and Human Factors* 1(1): 3–15.
- Mehta, R.K., Nussbaum, M.A. and Agnew, M.J. 2012. Muscle- and task-dependent responses to concurrent physical and mental workload during intermittent static work. *Ergonomics* 59(10): 1166–1179.
- Mehta, R.K. and Parasuraman, R. 2014. The effect of mental fatigue on the development of physical fatigue: A neuroergonomic approach. *Human Factors*, 56(4): 645–656.
- Navon, D. 1984. Resources – A theoretical soup stone? *Psychological Review* 91(2): 216.
- Nelson, J.T., McKinley, R.A., Golob, E.J., Warm, J.S. and Parasuraman, R. 2014. Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tDCS). *NeuroImage* 85, 907–917. dx.doi.org/10.1016/j.neuroimage.2012.11.061.
- Nuechterlein, K.H., Parasuraman, R. and Jiang, Q. 1983. Visual sustained attention: Image degradation produces rapid sensitivity decrement over time. *Science* 220(4594): 327–329.
- Nybo, L. and Rasmussen, P. 2007. Inadequate cerebral oxygen delivery and central fatigue during strenuous exercise. *Exercise and Sport Sciences Reviews* 35(3): 110.
- Ogoh, S., Sato, K., Akimoto, T. et al. 2010. Dynamic cerebral autoregulation during and after handgrip exercise in humans. *Journal of Applied Physiology* 108(6): 1701–1705.
- Omlor, W., Patino, L., Hepp-Reymond, M.C. and Kristeva, R. 2007. Gamma-range corticomuscular coherence during dynamic force output. *NeuroImage* 34(3): 1191–1198.
- Parasuraman, R. 1987. Human-computer monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 29(6): 695–706.
- Parasuraman, R. 2000. Designing automation for human use: Empirical studies and quantitative models. *Ergonomics* 43(7): 931–951.
- Parasuraman, R. 2003. Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science* 4(1–2): 5–20.
- Parasuraman, R. 2011. Neuroergonomics brain, cognition, and performance at work. *Current Directions in Psychological Science* 20(3): 181–186.
- Parasuraman, R., Bahri, T., Deaton, J.E., Morrison, J.G. and Barnes, M. 1992. *Theory and Design of Adaptive Automation in Aviation Systems*. Warminster, PA: Naval Air Development Center.
- Parasuraman, R. and Caggiano, D. 2005. Neural and genetic assays of human mental workload. *Quantifying Human Information Processing*, eds., D. McBride and D. Schmorrow, pp. 123–149. Oxford: Rowman and Littlefield.
- Parasuraman, R., de Visser, E., Clarke, E. et al. 2009. Detecting threat-related intentional actions of others: Effects of image quality, response mode, and target cuing on vigilance. *Journal of Experimental Psychology: Applied* 15(4): 275.
- Parasuraman, R. and Rizzo, M. 2007. *Neuroergonomics: The Brain at Work*, Vol. 195177614. Oxford, U.K.: Oxford University Press.
- Parasuraman, R., Warm, J.S. and See, J.E. 1998. Brain systems of vigilance. In *The Attentive Brain*, ed. R. Parasuraman, pp. 221–256. Cambridge, MA: MIT Press.
- Parasuraman, R. and Wilson, G.F. 2008. Putting the brain to work: Neuroergonomics past, present, and future. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50(3): 468–474.
- Perrey, S. 2008. Non-invasive NIR spectroscopy of human brain function during exercise. *Methods* 45(4): 289–299.
- Pike, M.F., Maior, H.A., Porcheron, M., Sharples, S.C. and Wilson, M.L. 2014. Measuring the effect of think aloud protocols on workload using fNIRS. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pp. 3807–3816. Washington, DC: ACM.
- Pizzagalli, D.A. 2007. Electroencephalography and high-density electrophysiological source localization. In *Handbook of Psychophysiology*, eds. J.T. Cacioppo, L.G. Tassinari and G.G. Bernston, 3rd edn., pp. 56–84. Cambridge, U.K.: Cambridge University Press.
- Plichta, M., Herrmann, M., Baehne, C. et al. 2006. Event-related functional near-infrared spectroscopy (fNIRS): Are the measurements reliable? *NeuroImage* 31(1): 116–124.

- Poldrack, R., Mumford, J.A. and Nichols, T.E. 2011. *Handbook of Functional MRI Analysis*. New York: Cambridge University Press.
- Posner, M.I. 2012. Expanding horizons in ergonomics research. *NeuroImage* 59: 149–153.
- Rowe, J., Toni, I., Josephs, O., Frackowiak, R. and Passingham, R. 2000. The prefrontal cortex: Response selection or maintenance within working memory? *Science* 288(5471): 1656.
- Satterfield, K., Ramirez, R., Shaw, T. and Parasuraman, R. 2012. Measuring workload during a dynamic supervisory control task using cerebral blood flow velocity and the NASA-TLX. In *Proceedings of the Human Factors and Ergonomics Society*, Boston, MA.
- Scerbo, M. 2007. Adaptive automation. In *Neuroergonomics: The Brain at Work*, eds. R. Parasuraman and M. Rizzo, pp. 239–252. New York: Oxford.
- Scholz, J., Klein, M.C., Behrens, T.E.J. and Johansen-Berg, H. 2009. Training induces changes in white-matter architecture. *Nature Neuroscience* 12(11): 1370–1371.
- Schreppel, T., Egetemeir, J., Schecklmann, M. et al. 2008. Activation of the prefrontal cortex in working memory and interference resolution processes assessed with near-infrared spectroscopy. *Neuropsychobiology* 57(4): 188–193.
- Schultheis, H. and Jamieson, A. 2004. Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. In *Adaptive Hypermedia and Adaptive Web-Based Systems*, eds. P.D. Bra and W. Nejdl, pp. 18–24. Eindhoven, the Netherlands: Springer.
- Sen, P.N. and Basser, P.J. 2005. A model for diffusion in white matter in the brain. *Biophysical Journal* 89(5): 2927–2938.
- Shaw, T.H., Warm, J.S., Finomore, V. et al. 2009. Effects of sensory modality on cerebral blood flow velocity during vigilance. *Neuroscience Letters* 461(3): 207–211.
- Shibuya, K. and Kuboyama, N. 2007. Human motor cortex oxygenation during exhaustive pinching task. *Brain Research* 1156: 120–124.
- Shortz, A., Van Dyke, S. and Mehta, R.K. 2012. Neural correlated of physical and mental fatigue. In *Proceedings of the Human Factors and Ergonomics Society*, Boston, MA.
- Siemionow, V., Yue, G.H., Ranganathan, V.K., Liu, J.Z. and Sahgal, V. 2000. Relationship between motor activity-related cortical potential and voluntary muscle activation. *Experimental Brain Research* 133(3): 303–311.
- Spiers, H.J. and Maguire, E.A. 2007. Decoding human brain activity during real-world experiences. *Trends in Cognitive Sciences* 11(8): 356–365.
- Stammers, R.B. and Patrick, J. 1975. *Psychology of Training*. London, U.K.: Methuen.
- Stroobant, N. and Vingerhoets, G. 2000. Transcranial Doppler ultrasonography monitoring of cerebral hemodynamics during performance of cognitive tasks: A review. *Neuropsychology Review* 10(4): 213–231.
- Takeuchi, H., Sekiguchi, A., Taki, Y. et al. 2010. Training of working memory impacts structural connectivity. *The Journal of Neuroscience* 30(9): 3297–3303.
- Taylor, J.L. and Gandevia, S.C. 2008. A comparison of central aspects of fatigue in submaximal and maximal voluntary contractions. *Journal of Applied Physiology* 104(2): 542–550.
- Thomas, R. and Stephane, P. 2008. Prefrontal cortex oxygenation and neuromuscular responses to exhaustive exercise. *European Journal of Applied Physiology* 102(2): 153–163.
- Ting, C.H., Mahfouf, M., Nassef, A. et al. 2010. Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 40(2): 251–262.
- Tomprowski, P.D. 2003. Effects of acute bouts of exercise on cognition. *Acta Psychologica* 112: 297–324.
- Ushiyama, J., Katsu, M., Masakado, Y. et al. 2011. Muscle fatigue-induced enhancement of corticomuscular coherence following sustained submaximal isometric contraction of the tibialis anterior muscle. *Journal of Applied Physiology* 110(5): 1233–1240.
- Ushiyama, J., Takahashi, Y. and Ushiba, J. 2010. Muscle dependency of corticomuscular coherence in upper and lower limb muscles and training-related alterations in ballet dancers and weightlifters. *Journal of Applied Physiology* 109(4): 1086–1095.
- Utz, K.S., Dimova, V., Oppenländer, K. and Kerkhoff, G. 2010. Electrified minds: Transcranial direct current stimulation (tDCS) and galvanic vestibular stimulation (GVS) as methods of non-invasive brain stimulation in neuropsychology – A review of current data and future implications. *Neuropsychologia* 48(10): 2789–2810.
- Villringer, A. and Chance, B. 1997. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences* 20(10): 435–442.
- Walsh, V. and Pascual-Leone, A. 2005. *Transcranial Magnetic Stimulation: A Neurochronometrics of Mind*. Cambridge, MA: Bradford.

- Wang, L., Tan, T., Hu, W. and Ning, H. 2003. Automatic gait recognition based on statistical shape analysis. *IEEE Transactions on Image Processing* 12: 1120–1131.
- Wang, Z., Hope, R.M., Wang, Z., Ji, Q. and Gray, W.D. 2012. Cross-subject workload classification with a hierarchical Bayes model. *NeuroImage* 59(1): 64–69.
- Warm, J.S., Dember, W.N. and Hancock, P.A. 1996. Vigilance and workload in automated systems. In *Automation and Human Performance: Theory and Applications*, eds. R. Parasuraman and M. Mouloua, pp. 183–200. Mahwah, NJ: Erlbaum.
- Warm, J.S. and Parasuraman, R. 2007. Cerebral hemodynamics and vigilance. In *Neuroergonomics: The Brain at Work*, eds. R. Parasuraman and M. Rizzo, pp. 146–158. New York: Oxford University Press.
- Warm, J.S., Parasuraman, R. and Matthews, G. 2008. Vigilance requires hard mental work and is stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50(3): 433–441.
- Wickens, C.D. 1984. Processing resources in attention. In *Varieties of Attention*, eds. R. Parasuraman and R. Davies, pp. 63–101. Orlando, FL: Academic Press.
- Wickens, C.D. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3(2): 159–177.
- Wickens, C.D. and McCarley, J.S. 2008. *Applied attention theory*. Boca Raton, FL: CRC Press, Taylor & Francis Group.
- Wiener, E.L. and Attwood, D.A. 1968. Training for vigilance: Combined cueing and knowledge of results. *Journal of Applied Psychology* 52(6p1): 474.
- Wilson, G.F. 2001. In-flight psychophysiological monitoring. In *Progress in Ambulatory Monitoring*, eds. F. Fahrenberg and M. Myrtek, pp. 435–454. Seattle, WA: Hogrefe & Huber.
- Wilson, G.F. 2002. Psychophysiological test methods and procedures. In *Handbook of Human Factors Testing and Evaluation*, eds. S.G. Charlton and T.G. O'Brien, 2nd edn., pp. 127–156. Mahwah, NJ: Erlbaum.
- Wilson, G.F. and Russell, C.A. 2003. Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors* 45(3): 381–389.
- Wilson, G.F. and Russell, C.A. 2007. Performance enhancement in an uninhabited air vehicle task using psychophysiological determined adaptive aiding. *Human Factors* 49(6): 1005–1018.
- Yanagisawa, H., Dan, I., Tsuzuki, D. et al. 2010. Acute moderate exercise elicits increased dorsolateral prefrontal activation and improves cognitive performance with Stroop test. *NeuroImage* 50(4): 1702–1710.
- Yang, Q., Fang, Y., Sun, C.-K. et al. 2009. Weakening of functional corticomuscular coupling during muscle fatigue. *Brain Research* 1250(0): 101–112.
- Yoon, T., Keller, M.L., De-Lap, B.S. et al. 2009. Sex differences in response to cognitive stress during a fatiguing contraction. *Journal of Applied Physiology* 107(5): 1486.
- Yue, G.H., Liu, J.Z., Siemionow, V. et al. 2000. Brain activation during human finger extension and flexion movements. *Brain Research* 856(1–2): 291–300.

# *Section IV*

---

## *Assessment and Design of the Physical Workplace*

The physical work environment is something that is relevant to all contexts, and is one of the most widely explored of any groups of ergonomics variables. The chapters within this section outline the different types of method that the ergonomics/human factors (E/HF) practitioner is able to use to assess the thermal, visual and auditory environments and to use an understanding of anthropometry to inform the physical layout of a workplace.

Each method outlines the fundamental human characteristics or capabilities that must be taken into account when conducting assessments or making design interventions. In many ways, these are the ‘limiting characteristics’ of which any E/HF specialist must be aware; however, it is also clear in all of the chapters that there is considerable variation between both the perceptual and physical capabilities of individuals and their personal preferences for workplace design and configuration. This is perhaps counter-intuitive, and even frustrating; here we have a variable where we can clearly measure perceptual sensations and capabilities, but still, there are so many situational, experiential and attitudinal variables interacting, that we cannot easily make confident predictions about individuals’ comfort or capabilities. This therefore again emphasises the need to not only use prediction techniques, based on sound theoretical patterns, to design physical workplaces, but also to apply user-centred methods to check our designs and collect participant responses.

Of course, as outlined in Chapter 1, it is critical that the design of the working environment takes into account E/HF requirements as much as possible – if designed well, an environment will lead to better work culture and performance. In the case of the physical work place, this is often a case of reducing the presence of variables that lead to discomfort, such as flicker from lights, irritating air draughts, noise annoyance or cramped workspaces. Such variables can not only lead to lower employee satisfaction, but can also lead to reduced performance. This can occur because of health problems or discomfort and dissatisfaction, but may also be produced more directly. Examples are disability glare on a display, speech interference noise levels, cold conditions leading to loss of dexterity or reduced vigilance through distractions from poor seating. Possible performance decrement has been proposed in terms of output and errors, for mental and physical tasks, and with respect to both direct task-related measures and also systemic ones such as absenteeism or labour turnover. It must be said that, in general, results on performance effects of the environment are more equivocal than those on health or discomfort/dissatisfaction.

This section opens with two chapters that focus on the thermal environment. Parsons (Chapter 23) firstly presents an approach to be used to conduct an environmental survey and follows this (Chapter 24) with a detailed description of the theory explaining the relationship between thermal environment characteristics and the resulting human physiological and psychological response. Howarth (Chapter 25) considers the visual environment, explaining the way that light behaves in different situations and the way our visual system responds and adapts to the vastly different visual stimuli as we go about our work and lives. Haslegrave (Chapter 26) addresses the auditory environment, again describing the relationship between the physical characteristics of sound and the human response to auditory stimuli in different types of work situations. In each of these three chapters (24–26), the authors describe the units of measurement of environmental stimuli, the different human responses of relevance and the range of methods and techniques available for capturing the characteristics of the environment and individuals' response to that environment.

The final chapter in this section, by Feathers, Clive D'Souza and Victor Paquet, explains the relationship between human anthropometric characteristics and the design of workplaces. They outline the different sources of data available to the E/HF practitioner and describe how anthropometric data and calculations can be applied to different types of workplace layout and design activities. Superficially, the procedures are simple, but in fact, there are many poor applications of anthropometry, which arise to a great extent through insufficient recognition of the reasons for the choice of the various dimensions or percentiles used.

It is perhaps the case that the area of physical workplace design has not received as much active research attention as the areas of organisational or cognitive E/HF in recent years, but this does not mean that this type of E/HF has been 'done'. Our workplaces continue to change, and as we note elsewhere in this book, there are many jobs now where a user is mobile, uses multiple work locations and technologies and blurs their work–life activities. There is also some evidence that people are also evolving in responses to these changes, for example, in studies that have examined the impact of smartphone or games technology use on physical capabilities. There is also the need to ensure that any physical workplace design accommodates as wide a population as possible; in addition to adopting inclusive design principles (as described in Chapter 11), it is important to acknowledge the aging (and increasingly international) nature of both the work population and the population as a whole. There is therefore a particular need to ensure that the data that underpin our theoretical models of response to work environments are as applicable as possible to the working population now and the wider population of future work and life.

**Sarah Sharples**



---

# 23 The Environmental Ergonomics Survey

*Ken Parsons*

## CONTENTS

|                                                                                                                                                                                                                    |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Introduction.....                                                                                                                                                                                                  | 641 |
| Environmental Ergonomics.....                                                                                                                                                                                      | 642 |
| Global Activity .....                                                                                                                                                                                              | 642 |
| Practical Effects of Environments on People.....                                                                                                                                                                   | 642 |
| Environment and People: Principles .....                                                                                                                                                                           | 643 |
| Environmental Ergonomics Methods.....                                                                                                                                                                              | 644 |
| Human Participant Trials .....                                                                                                                                                                                     | 645 |
| Specify the Aim.....                                                                                                                                                                                               | 645 |
| Which Participants and How Many?.....                                                                                                                                                                              | 645 |
| ISO 28802 ‘Ergonomics of the Physical Environment – The assessment of environments by means of an environmental survey involving physical measurements of the environment and subjective responses of people’..... | 646 |
| Case Study: Environmental Survey of an Office Where People Refused to Work .....                                                                                                                                   | 648 |
| The Problem.....                                                                                                                                                                                                   | 648 |
| The Environmental Ergonomics Survey .....                                                                                                                                                                          | 648 |
| The ‘Expert’ Checklist.....                                                                                                                                                                                        | 648 |
| Questionnaire .....                                                                                                                                                                                                | 648 |
| Interviews with Staff .....                                                                                                                                                                                        | 649 |
| Environmental Measurements.....                                                                                                                                                                                    | 649 |
| Wrap-Up Meeting .....                                                                                                                                                                                              | 651 |
| Report.....                                                                                                                                                                                                        | 652 |
| References.....                                                                                                                                                                                                    | 653 |

## INTRODUCTION

A frequent request to ergonomics and human factors (E/HF) practitioners is to conduct an environmental survey to determine why people are complaining, why the environment is considered uncomfortable and whether productivity and health are affected. It is also important to determine what we can do about it. As well as allowing an assessment of existing and established environments, an important application of a survey in newly constructed or occupied buildings will be to determine whether the buildings are built to specification in terms of environments for human occupation and whether the occupants themselves are satisfied with their new home, office, leisure area or other types of space. Thermal, acoustic, vibration, lighting and air quality environmental components and their interactions are included in the survey considered in this chapter as well as behavioural responses to the environment. ISO 28802 (2012) provides information of how to conduct an environmental ergonomics survey and is described.

This chapter provides practical advice on how to carry out a survey to determine the effects of the environment on people and make recommendations for improvement. Both general principles and practical examples are provided such that readers will be able to design and conduct an environmental survey of their own and have an understanding of the general issues. A more profound discussion of principles and practice in the areas of visual environment, thermal environment and acoustics is provided in Chapters 24 to 26. The environmental survey can be regarded as the first and often sufficient response to a practical question.

## ENVIRONMENTAL ERGONOMICS

Environmental ergonomics can be defined as the application of knowledge, of human response to the environment, to the design of systems. People in systems operate within an environment, and environmental ergonomics is concerned with how they interact with the environment from the perspective of the whole system. Although there have been many studies of human responses to the environment (light, noise, heat, cold, etc.) over hundreds of years and much is known, it is only with the development of ergonomics as a discipline that the unique features of environmental ergonomics are beginning to emerge. In principle, environmental ergonomics will encompass the social, psychological, cultural and organisational environments of systems; however, to date, it has been viewed as concerned with the individual components of the physical environment. Typically, ergonomists have considered the environment in a mechanistic way, in terms of the lighting or noise survey, for example, rather than an integral part of ergonomics investigation (e.g. if cold distracts the worker, what are the consequences for the overall system?).

## GLOBAL ACTIVITY

The establishment of the study of human responses to the physical environment has paradoxically inhibited the development of environmental ergonomics as it has produced associated institutions that provide inertia to the acceptance of an ergonomics approach. Examples include learned societies and conferences on specific aspects of the environment such as noise, lighting or vibration. For example, the *International Society for Environmental Ergonomics* first met in Bristol in 1984 and has since then held successful biannual conferences around the globe. The original intention was to provide a forum for environmental ergonomists in general; however, it very soon became specifically concerned with human responses to heat and cold. This provided a clear demonstration that there are few researchers and institutions that consider human responses to environments as a whole, rather than in terms of its component parts. The *International Standards Organisation* (ISO) and more recently *European Standards Organisation* (CEN) have made significant contributions in the area of environmental ergonomics. However, the existence of established standards committees in noise, vibration, lighting and others has hampered progress, as these bodies often take a product- or manufacturer-orientated perspective which is not human centred and not conducive to an integrated ergonomics approach. This position is not static however, and it has become increasingly recognised (e.g. ISO 28802, 2012; ISO 28803, 2012) that people experience total environments and that ergonomics methods are essential for effective practical application. Much knowledge exists, and new approaches will allow that knowledge to contribute to environmental ergonomics as a major and essential contribution to ergonomics investigation.

## PRACTICAL EFFECTS OF ENVIRONMENTS ON PEOPLE

There is a continuous and dynamic interaction between people and their surroundings that produces physiological and psychological strain on the person. This can lead to discomfort, annoyance, subtle and direct effects on performance and productivity, effects on health and safety, and death. Examples would include discomfort in offices due to glare, noisy equipment, draughts or smells.

In the cold, people experience frostbite and die from hypothermia. In the heat, they collapse or die from heat stroke. People exposed to vibrating tools have damage to their hands. Performance can be dramatically affected by loss of manual dexterity in the cold, noise interfering with speech communication or work time lost because the environment is unacceptable or distracting. Accidents can occur due to glare on displays, missed signals in a warm environment or disorientation due to exposure to extreme environments.

## ENVIRONMENT AND PEOPLE: PRINCIPLES

Most of the energy that makes up our environment originally comes as electromagnetic radiation from the sun. Around  $1373 \text{ W m}^{-2}$  (the solar constant) enters the outer limits of the earth's atmosphere, and this arrives on the earth in modified form where it is transformed from place to place and from one form to another (heat, mechanical, light, chemical, electrical). The wide diversity of environments to which people are exposed is therefore defined by that energy which varies in level characteristic and form. It is the human condition to interact and survive in those environments and part of that has been the creation of 'local' optimum environments, for example, buildings.

The human body is not a passive system that responds to an environmental input in a way that is monotonically related to the level of the physical stimulus. Any response depends upon a great number of factors. If viewed in engineering terms, the 'transducers' of the body (sensors – eyes, ears, etc.) have their own specification in terms of responses to different types of physical stimuli (e.g. the eyes have spectral sensitivity characteristics). In addition, the body does not behave as a passive system; for example, the body responds to a change in environmental temperature by reacting in a way consistent with maintaining internal body temperature (e.g. by sweating to lose heat by evaporation). The body therefore senses the environment with a 'transducer' system that has its own characteristics, and it reacts in a dynamic way to environmental stimuli.

The aforementioned engineering model is simplistic. There are many other factors involved. For example, the way in which a stimulus is perceived and hence any response to it will depend upon that person's past experience, their emotional state at the time and other factors. It is with consideration of these physical, physiological and psychological factors that the environmental ergonomist must provide a practical solution to the problems of how a human occupant will respond to an environment.

A paradigm shift in designing environments for people is to consider 'adaptive opportunity'. This refers to the opportunity and ability of the people in the environment to change the environment to which they are exposed. For example, work, tasks and environments can be designed to provide the opportunity for people to move around, change posture, adjust clothing, adjust temperature or lighting, open and close windows and so on. The task of the environmental designer is then not to design to specification but to provide a range of conditions within which people can adapt or behave in order to achieve the optimum environment for them. This can also allow for energy savings and cost in producing the environments and for people in different cultures to achieve 'comfort' in their own way (Parsons 2014). An environmental survey, therefore, will need to consider the adaptive opportunities provided by the environment and how people are using them.

An additional factor that must be considered is that of individual differences. These can be conveniently divided into inter-individual differences that are differences between people (e.g. males and females, tall and short people) and intra-individual differences that are differences that occur in the same person over time (e.g. emotional state, menstrual cycle changes in females). There are ways in which design can be made for specific individuals. However, it is usual in practice to design for a population of users. It is often adequate therefore to describe individual differences in terms of statistical parameters of the population (e.g. mean and standard deviation of responses – see Chapter 2).

Knowledge of individual differences in responses to environments has increased over the years, and it is now possible to consider specific individuals and populations in environmental design.

Environments for people with disabilities, cultural differences, gender differences and possible differences in types of preferred environments across the world, because people have different behaviours, expectancies and preferred conditions, are all important. A successful survey will have to take these into account. In conducting the survey, use of language as well as the social, psychological and cultural context of the survey must be considered.

## ENVIRONMENTAL ERGONOMICS METHODS

There are four principal methods of assessing human response to environments: subjective methods, where those representative of the user population actually report on the response to the environment; objective measures, where the occupant's response is directly measured (e.g. body temperature, hearing ability, performance at a task); behavioural methods, where the behaviour of a person or group is observed and related to responses to the environment (e.g. change posture, move away, switch on lights); and modelling methods. Modelling methods include those where predictions of human response are made from models that are based on experience of human response in previously investigated environments (empirical models) or rational models of human response to environments that attempt to simulate the underlying system and hence can be used to relate cause and effect.

Subjective methods include the use of simple rating scales, of thermal comfort, and also more detailed responses and questionnaires; they could also include discourse analysis and focus groups. They have the advantage of being relatively easy to carry out and are particularly suited to assessing psychological responses such as comfort and annoyance. They can also usefully be used when the contributing factors to a response are not known. They have the disadvantage of being difficult to design having a number of potential methodological biases (see Chapter 4). In addition, subjective methods are often not appropriate for assessing such things as effects on health; for example, a person cannot always detect when they are under physiological strain. They may also not be appropriate for use with young children or for people with cognitive impairments. Also an environmental stress can interfere with a person's capacity to make a reliable subjective assessment. A further disadvantage is that subjective methods often require the use of a representative sample of the user population being exposed to the environment of interest. This is cumbersome if used in initial design. From a global perspective, if standard subjective scales are used, then it is important that translations of language (e.g. from English) reflect the meaning of the scale to the population of interest. It cannot be assumed that people across the world will have the same attitude towards subjective assessments as well as the same interpretation of what is being requested. In some cultures, there may be some reluctance to express dissatisfaction, (Tochihara et al., 2012).

Objective methods have the advantage of providing direct measures of human response. This could include measures of body temperature and transmitted acceleration to the head from vibration inputs, as well as direct measures of performance at a task. The main disadvantages are that a representative sample of the user population is required to be exposed to the environment of interest (not useful for design), the measuring instruments can interfere with what they are used to measure and objective measures cannot easily predict states such as comfort.

Behavioural methods are probably underused in environmental ergonomics. They can have the unique advantage of not interfering with what they are attempting to measure. They include changes in posture, changing clothing, adjusting the environment, moving away and working faster or slower. A 'model' is needed to interpret the 'reason' for any behaviour. Observer training is required. Improvements in behavioural modelling and remote sensing technology provide opportunities for the future as environmental ergonomics makes more use of behavioural techniques in the context of considering the environment as an integral part of complex systems. These methods are particularly suited for studying some people with disabilities, children or other special populations, or contexts where other methods would be inappropriate. A difficulty is determining cause

and effect. Did the person change posture because they were too hot or was the chair uncomfortable or the line of sight obscured? So-called adaptive opportunity provides an indication of the possibility for people to adapt (move away, adjust clothing, etc.). An otherwise unacceptable environment may become acceptable if it is designed with adaptive opportunity.

Models of human response to environments have the advantage of being consistent in their response, are easy to use, give a quick response and can be used in both design and evaluation. The main disadvantages are that the models provide only approximate responses when designing for individuals, and inevitably, there will be factors in any real environment which the models will not consider.

The method that is most appropriate will depend upon the specific investigation under consideration. A simple environmental survey may take a 'checklist' approach, where the ergonomist completes a list of questions about the environment (see Chapter 4). This may be complemented by interviews or stand-alone questionnaires completed by people who occupy the environment and the use of simple measuring instruments to provide a first indication of the nature of the environment (e.g. temperature, humidity, noise levels, light levels).

If the environment is new or a specific question is being asked (e.g. comparison of two rooms, assessing different lighting systems), then a more formal approach using human participants may be required. The following sections present practical advice on how to conduct trials with human participants and a practical assessment of a room where workers have complained about their environment.

## **HUMAN PARTICIPANT TRIALS**

The design of any test or trial using human participants will depend upon the specific aims of that test or trial. However, there are general principles and these are outlined in the following text. A typical trial involves exposing people to environments of interest in a controlled way, measuring environmental conditions and recording the responses of people.

### **SPECIFY THE AIM**

An optimum trial design will achieve its aim with efficient use of resources. To achieve this, it is important to be clear about the specific aim or aims. For example, if the aim is to compare three types of environments, then a repeated measures design, where all participants are exposed to all environments (in a balanced order) in identical conditions, may provide the best comparison. Contrast this with the evaluation of an index where a wide range of environmental conditions may be optimum. If both aims need to be met, then both must be met in the design. It is necessary therefore to be specific about the aims of the trial.

### **WHICH PARTICIPANTS AND HOW MANY?**

A valid method of evaluating environments would be to use a panel of experts. This technique is used in wine tasting, for example, where acknowledged experts give opinions concerning the quality of wines. This technique depends upon identifying unbiased acknowledged experts. This is usually not possible in the area of environmental assessment and the trial designs should specifically avoid bias. It is usual to identify a 'random' sample of participants as representatives of the population of interest. This is a question of statistical sampling, and relevant factors such as age, gender, experience and anthropometry could be identified and influence selection. The number selected will depend upon the aim and experimental design. A calculation can be made based upon the power of a statistical test, that is, the probability of accepting the alternative hypothesis (e.g. environment A is more comfortable than environment B) given that it is true. This is a rather academic approach and requires assumptions to be made about the strength of effect you expect, which is rather circular

as this is what you are trying to find out (see Chapter 2 for further discussion of power analysis in studies). Of practical importance will be the allocation of participants to treatments. If there are three rooms and three types of glazing being compared (i.e. nine conditions), then nine participants would allow a  $9 \times 9$  Latin square design. That is where each participant is exposed to each condition in a different, balanced, order. A repeated measures design is where all participants are exposed to all conditions (see Chapter 2).

Although not statistically rigorous, other pointers are useful. It is generally considered that, for normally distributed responses, increasing the number of participants provides a diminishing return in terms of a sample representing a population. Numbers greater than eight are often considered an acceptable sample size. It is also useful to consider approximate probability. For example, if two vehicles were compared by four people, then the probability of all four preferring vehicle A to vehicle B due to chance (when there is actually no difference in comfort between the vehicles) is  $1/2$  to the power of 4 =  $1/16 = 6.25\%$ . So a sample of four would not be sufficient to make a decision at a 5% level even in the case of an extreme result. An example of practical significance is whether the experimenter would be satisfied that if all their participants preferred A to B, then this is considered sufficient evidence that A is more comfortable than B. It is useful therefore to estimate how many participants it would take for practical significance to be established. It may be that statistical significance may be established with the use of large groups, but the effect may be small and not of practical significance. A more rigorous statistical approach can be taken in any particular experiment; however, the aforementioned 'rules of thumb' can be useful. The use of experimental trials in a formal assessment of a specific hypothesis is particularly useful in environmental design (e.g. which heating systems to select). More detailed information on the use of human participants is covered comprehensively in Chapters 1 to 8.

## **ISO 28802 'ERGONOMICS OF THE PHYSICAL ENVIRONMENT – THE ASSESSMENT OF ENVIRONMENTS BY MEANS OF AN ENVIRONMENTAL SURVEY INVOLVING PHYSICAL MEASUREMENTS OF THE ENVIRONMENT AND SUBJECTIVE RESPONSES OF PEOPLE'**

There is international recognition that there is often a requirement to provide an assessment of an integrated environment such as an office, building, vehicle and outside space. It could be used in an assessment where complaints had been received and workers are unhappy, as part of a post-occupancy questionnaire or for other applications. A first approach to assessment is to conduct an environmental survey, and this often involves physical measures of the environment and measurements of the responses of the people who occupy the space being surveyed. Physical measures can be compared with recommended values and subjective responses will indicate actual levels of comfort and satisfaction.

ISO 28802 (2012) is an international standard that provides guidance on conducting an environmental survey. It is systematic in its approach, considering each environmental component in turn, and provides a powerful tool in assessing integrated environments which will lead to a report and recommendations for improvement. Table 23.1 provides a summary of the standard, describing which parameters to measure (e.g. horizontal illuminance; see Chapter 25), instruments for measurement and suggested subjective scales for each environmental component.

Appendix 1 of the standard provides a simple observation checklist that should be completed by the person doing the survey which will complement the measurements taken in the survey. The environmental survey presented in the standard provides a comprehensive first assessment that may lead to further, more detailed, analysis in areas identified for improvement. As an international standard, it provides support for use of the environmental survey as a procedure that will be used across the world.



**TABLE 23.1**

**Parameters, Measurement, Instrumentation and Suggested Subjective Scales for Use in the Environmental Survey Proposed in ISO 28802 (2012)**

| Environment         | Parameters                                                                                                                    | Instrument Specification                 | Subjective Terms              | Suggested Subjective Scales (sl. – Slightly)                                                                                |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|-------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Thermal             | Air temperature                                                                                                               | ISO 7726                                 | ISO 10551                     |                                                                                                                             |
|                     | Radiant temperature                                                                                                           | ISO 7726                                 | Sensation 3 to −3             | Hot, warm, sl. warm, neutral, sl. cool, cool, cold                                                                          |
|                     | Humidity                                                                                                                      | ISO 7726                                 | Uncomfortable 1–4             | Not uncomfortable, sl. uncomfortable, uncomfortable, very uncomfortable                                                     |
|                     | Air velocity                                                                                                                  | ISO 7726                                 | Stickiness 1–4                | Not sticky, sl. sticky, sticky, very sticky                                                                                 |
|                     | Metabolic rate                                                                                                                | ISO 8996                                 | Preference 1–7                | Much warmer, warmer, sl. warmer, no change, sl. cooler, cooler, much cooler                                                 |
|                     | Clothing insulation                                                                                                           | ISO 9920<br>ISO 7730                     | Draughtiness 1–4              | Not draughty, sl. draughty, draughty, very draughty                                                                         |
|                     |                                                                                                                               |                                          | Dryness 1–4                   | Not dry, sl. dry, dry, very dry                                                                                             |
| Satisfaction 1–2    |                                                                                                                               |                                          | Satisfied, not satisfied      |                                                                                                                             |
| Acceptability 1–2   |                                                                                                                               |                                          | Acceptable, not acceptable    |                                                                                                                             |
| Acoustical          | dB(A)                                                                                                                         | IEC 61672-1                              | Annoyance 1–4                 | Not annoying, sl. annoying, annoying, very annoying                                                                         |
|                     | dB(A) Leq.                                                                                                                    | IEC 61672-1                              | Preference 1–4                | No change, sl. quieter, quieter, much quieter                                                                               |
|                     |                                                                                                                               | ISO 9612                                 | Acceptability 1–2             | Acceptable, not acceptable                                                                                                  |
| Visual and lighting | Horizontal illuminance                                                                                                        | CIE 69 standard<br><br>Illuminance meter | Satisfaction 1–2              | Satisfied, not satisfied                                                                                                    |
|                     |                                                                                                                               |                                          | Visual discomfort 1–4         | No discomfort, sl. discomfort, discomfort, much discomfort                                                                  |
|                     |                                                                                                                               |                                          | Preference 1–7                | Much darker, darker, sl. darker, no change, sl. lighter, lighter, much lighter                                              |
|                     |                                                                                                                               |                                          | Acceptability 1–2             | Acceptable, not acceptable                                                                                                  |
| Air quality         | CO <sub>2</sub> levels                                                                                                        | Specified by manufacturers               | Satisfaction 1–2              | Satisfied, not satisfied                                                                                                    |
|                     |                                                                                                                               |                                          | Smelliness                    | Very smelly, smelly, sl. smelly, not smelly                                                                                 |
|                     |                                                                                                                               |                                          | Acceptability 1–2             | Acceptable, not acceptable                                                                                                  |
| Vibration           | Acceleration in vertical, horizontal and fore-and-aft directions with respect to a person. Sometimes also roll, pitch and yaw | ISO 8041<br><br>ISO 2631-1               | Satisfaction 1–2              | Satisfied, not satisfied                                                                                                    |
|                     |                                                                                                                               |                                          | Uncomfortable 1–6             | Not uncomfortable, a little uncomfortable, fairly uncomfortable, uncomfortable, very uncomfortable, extremely uncomfortable |
|                     | Frequency-weighted rms or vibration dose value                                                                                |                                          | Annoyance 1–4                 | Not annoying, sl. annoying, annoying, very annoying                                                                         |
|                     |                                                                                                                               |                                          | Acceptability<br>Satisfaction | Acceptable, not acceptable<br>Satisfied, not satisfied                                                                      |

## **CASE STUDY: ENVIRONMENTAL SURVEY OF AN OFFICE WHERE PEOPLE REFUSED TO WORK**

The following is a practical example of an environmental survey which is an actual case. It describes what was done and what was found and includes actual materials used.

### **THE PROBLEM**

The ergonomist received a request for assistance from managers of an office complex where in a particular large open plan office, workers were refusing to work and some were refusing to enter the room. The request arrived via email with attached photographs of the buildings and office. The workers complained of feeling unwell with unusual odours, tastes, facial tensions, headaches and lethargy. Medical opinion had recommended tests for carbon monoxide build up which had proven negative. Because of the seriousness of the case, the manager had been directed to seek expert assistance. Did they have a sick building?

### **THE ENVIRONMENTAL ERGONOMICS SURVEY**

It was agreed that the ergonomist would make a 'first shot' environmental ergonomics survey and make recommendations for a solution and for further action. The ergonomist conducted a 1-day assessment that included arrival and briefing by the managers, observation and measurements in the office, checklist completion by the ergonomist, completion of stand-alone questionnaires by the staff, interview of individual workers and debriefing of managers. A report and recommendations were provided soon after the assessment.

### **THE 'EXPERT' CHECKLIST**

The expert checklist used by the ergonomist is shown in Figure 23.1. The checklist is designed to 'force' the expert to systematically address important questions and to provide direct views concerning what is significant (e.g. What is the 'best' aspect?).

It can be seen that the office appeared well equipped with a good layout and modern workstations. The fabric of the building was poor and there was a fusty smell. The best aspect was the modern attractive equipment with good layout and that the room was quiet. The worst aspect was that there was poor building fabric and a smell. Air quality seemed poor with some damp. The thermal environment was acceptable, although a supplementary heater was noted. Daylight was present, but when dull outside, the lights provided poor colour rendering and the strip light was not working. It was a quiet office with good appearance and possibility to open windows. Although caution must be taken concerning 'expert' subjective opinion, it was evident that air quality was a problem.

### **QUESTIONNAIRE**

A simple questionnaire (Figure 23.2) was given to each of the workers at their workplaces (although some workers refused to enter the room and one of the [female] workers was immediately ill on entering and removed themselves, providing a modification to the context in which the questionnaire was completed). The workforce answered questions on how they generally felt, and how they felt now.

Overall the workers were generally satisfied with their thermal, lighting and acoustic environments. They found the room very smelly; they were not satisfied with the air quality and identified

### PHYSICAL ENVIRONMENT AND HUMAN PERFORMANCE CHECKLIST

Complete this checklist in the context of the organisational culture and mission, the job requirements of the staff and how the physical environment may affect their performance and productivity.

General impression (one sentence and one word descriptors)

Good points (include what is the best aspect)

Bad points (include what is the worst aspect)

Air quality (immediate impression on entering – stuffy, smelly, dusty?)

Thermal environment (hot, cold, humid, draughty, hot/cold surfaces, sweaty)

Lighting and visual environment (easy to see, lighting levels, clean windows and lights, glare, general appearance)

Noise and vibration (detect vibration, footfall, background noise level, interruptions, interference with task, annoying, noise sources)

Furniture (appearance and condition, fit for purpose, fit to persons size, telephone, chair)

Computer equipment (correctly positioned and adjusted, glare on screen, reflections, orientation)

Overall layout (storage space, organisation of work, filing system, coats and accessories)

Adaptive opportunity (clothing adjustment, move around, open window, control over conditions, level of activity, take breaks)

Distraction (sources of distraction from task that cease work or interfere with performance)

Overall conclusion : Environment optimum for performance?      Yes or No.

Recommendations:

**FIGURE 23.1** Environmental ergonomics ‘expert’ checklist to be completed by the investigator.

‘musty, damp’ smells and chemical smells. Some workers reported feeling dizzy, headaches and tiredness with irritation to the eyes and throat. Symptoms persisted when away from work and led to absenteeism.

One of the workers had been greatly troubled by the problem and had kept a diary of her experiences while working in the room. This included periods of illness attributed to working in the office.

### INTERVIEWS WITH STAFF

Individual interviews with staff confirmed the subjective reports. Some staff had been moved to the new office from what had previously been a better social environment. The problem had been going for a number of months and unions and medical personnel had been involved. Some staff felt that they had been identified as ‘complaining about nothing’. The interview detected a genuine interest in solving the problem and a genuine level of illness and frustration.

### ENVIRONMENTAL MEASUREMENTS

The air temperature and humidity were measured at each workplace with a whirling hygrometer. The air movement was observed by blowing (children’s) bubbles at the workplaces. Horizontal



**Noise**

9. Please indicate on the following scale how **YOU** find the **NOISE** in your environment **NOW**

|                   |       |
|-------------------|-------|
| Very annoying     | _____ |
| Annoying          | _____ |
| Slightly annoying | _____ |
| Not annoying      | _____ |

10. Please indicate any particular sources of **NOISE** that **YOU** can hear in your environment **NOW**.

11. Are you generally satisfied with the **NOISE** level in your environment at work?

Yes

No

12. Please give any additional information or comments which you think are relevant to the assessment of the **NOISE** in your environment at work (e.g. machines, talking, outside noise, etc.).

**Air Quality**

13. Please indicate on the following scale how **YOU** find the **AIR QUALITY** in your environment **NOW**

|                 |       |
|-----------------|-------|
| Very smelly     | _____ |
| Smelly          | _____ |
| Slightly smelly | _____ |
| Not smelly      | _____ |

14. Please indicate any particular sources of pollution that contribute to the **AIR QUALITY** in your environment **NOW**.

15. Are you generally satisfied with the **AIR QUALITY** in your environment at work?

Yes

No

16. Please give any additional information or comments which **YOU** think are relevant to the assessment of the **AIR QUALITY** in your environment at work (e.g. smells from smoke, chemicals, machines, etc.)

**General**

17. Do you suffer from persistent symptoms at work such as dry lips, eyes and throat, runny nose, etc.

Yes

No

18. Are you generally satisfied with your environment at work?

Yes

No

19. Please give any additional information or comments which **YOU** think are relevant to the assessment of your environment at work including suggestions for improvement.

**FIGURE 23.2 (Continued)** Environmental survey to be completed by workers at their workplace.

**WRAP-UP MEETING**

As the problem appeared to be one of air quality, an inspection of the building and building services was made with the Health and Safety and Building Services managers. A final discussion was then held with the managers. They were not clear if they had an environmental problem or an organisational problem. In any event the situation had persisted sufficiently for a managerial solution to be essential. Calling in experts would have helped the situation, and the assessment was sufficiently clear to conclude that there was an air quality issue.

It was noted that indoor air quality (IAQ) was a recognised design requirement for healthy buildings and is a global problem. The managers were referred to the *American Society of Heating, Refrigerating and Air-Conditioning Engineers* (ASHRAE) activity in this area. They defined IAQ as follows:

*.....the term indoor air quality (IAQ) represents the indoor air concentrations of pollutants that are known or suspected to affect people's comfort, environmental satisfaction, health, or work or school performance. Although this position document does not address thermal conditions, they are important for IAQ because temperature and humidity can affect pollutant emission rates, the growth of microorganisms on building surfaces, the survival of airborne infectious pathogens, the survival of house dust mites which are a source of allergens, people's perception of the quality of indoor air, prevalence rates of building related health symptoms, and work performance.*

Other indoor environmental factors such as, noise and lighting are also not addressed. Important references for more information on thermal conditions and other indoor environmental factors include ANSI/ASHRAE Standard 55–2007, Thermal Environmental Conditions for Human Occupancy, 'Indoor Air Quality Guide – Best Practices for Design, Construction and Commissioning', Chapter 9, Thermal Comfort, and Chapter 10, Indoor Environmental Health, of the. ASHRAE Handbook – Fundamentals, and ASHRAE Guideline 10, Interactions Affecting the Achievement of an Acceptable Indoor Environment.

The European standard EN 15251 (2007) 'Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics' was also noted.

## REPORT

A full environmental survey was not necessary as the problem was focussed around air quality. Lighting maintenance and improved colour rendering would provide some improvement. Three areas of advice were given for moving towards a solution. These are presented in the following in the actual letter sent to the managers from the ergonomist.

*To: Health and Safety Manager*

*John,*

### ENVIRONMENTAL ERGONOMICS SURVEY OF LARGE OPEN PLAN OFFICE

*Very pleased to meet with you yesterday. I present below a brief report of my findings. As you know I was in the building for only one day and made only a preliminary assessment. However I felt that the problem was identified with reasonable confidence that air quality was the issue and not other environmental components. Although there are always organisational issues associated with this sort of problem, I did not get the impression that they were severe in this case—for example in some organisations it is clear that the workers are not on board with the aims of the organisation and are not interested in solving the problems. In your case my impression was that there was a genuine interest and concern and it is worth keeping the workers on board.*

*As briefly discussed in our wrap-up meeting with Ted, my view is that there are environmental ergonomics, occupational hygiene, building services and management issues. I would suggest that all of these need to be addressed.*

1. *Environmental assessment – you could sample the air in the room and identify exactly what is in it. You could then compare the results with the Threshold Limit Values (TLVs) and other limits. The environmental air can include chemicals, biological contaminants*



- and particulates. I would recommend taking advice from a certified laboratory/expert consultant in this area. If you can find an experienced consultant he or she may be able to put their finger straight on the problem.*
2. *Building services – you should try and identify the source(s) of the problem. The source did not appear to be in the room although there was damp at the windows. The carpet is often the culprit but it seemed new. Does it get steam cleaned and are chemicals used? May be useful to have a word with the cleaners. Subjectively that whole area of the building did seem to exhibit odour however it was very much concentrated in the room. It is interesting that the room was the last on a lower level in the corridor. There seemed to be sources of chemicals in the building, from the toilets for example and an analysis of air flows within the building would be of interest. You could take expert advice on this and also make your own survey.*
  3. *Management issues – The implementation of the above investigations will be useful in showing that the problem is being taken seriously. The workers involved clearly have a problem and a management solution is inevitable. We discussed this at our meeting and there are a range of possibilities.*

#### *Who to contact*

*I will make some investigations into who to contact. However, it would be useful for you to contact the British Occupational Hygiene Society (BOHS) and the Chartered Institute of Building Services Engineers (CIBSE) for advice and a list of certified consultants. If you do contact a consultant be very clear about what you want otherwise you may get a specialist analysis of part of the problem. You want a whole solution. You also do not really want someone to say you have a problem/sick building etc. We have moved on from there.*

*I hope that the above is of some use.*

The aforementioned example describes an actual case where air quality was a problem. The checklist and questionnaire approach would also have identified other problems in the environment had they occurred. Simple on-the-spot solutions may then have been possible. If solutions were not obvious then for thermal environments, noise, vibration and lighting, a more detailed investigation could be recommended based upon the guidance provided in Chapters 24 to 26.

## REFERENCES

- EN 15251. (2007) Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics, CEN, Brussels.
- ISO 28802. (2012) Ergonomics of the physical environment – Assessment of environments by means of an environmental survey involving physical measurements of the environment and subjective responses of people, ISO, Geneva, Switzerland.
- ISO 28803. (2012) Ergonomics of the physical environment – Application of International Standards to people with special requirements, ISO, Geneva, Switzerland.
- Parsons K. C. (2014) *Human Thermal Environments*, CRC Press, Boca Raton, FL. ISBN 978-1-4665-9599-6.
- Tochihara Y., Lee J.-Y., Wakabayashi H., Wijayanto T., Bakri I., and Parsons K. C. (2012) The use of language to express thermal sensation suggests heat acclimatization by Indonesian people. *Int J Biometeorol*, 56(6): 1055–1064. doi: 10.1007/s00484–011- 0519-1.



---

# 24 Ergonomics Assessment of Thermal Environments

*Ken Parsons*

## CONTENTS

|                                                    |     |
|----------------------------------------------------|-----|
| Introduction.....                                  | 656 |
| The Principles .....                               | 656 |
| Relevant Measures .....                            | 656 |
| Thermoregulation.....                              | 657 |
| Thermal Indices.....                               | 657 |
| Heat Balance .....                                 | 657 |
| Rational Indices.....                              | 658 |
| Empirical Indices .....                            | 658 |
| Direct Indices .....                               | 658 |
| Measuring Instruments.....                         | 659 |
| Subjective Methods .....                           | 659 |
| Behavioural and Adaptive Methods .....             | 660 |
| Human Performance.....                             | 660 |
| Clothing.....                                      | 660 |
| Safe Surface Temperatures .....                    | 661 |
| International Standards .....                      | 663 |
| ISO Standards.....                                 | 663 |
| Thermal Models .....                               | 664 |
| The Practice .....                                 | 665 |
| Practical Assessment of Hot Environments .....     | 665 |
| Practical Assessment of Cold Environments .....    | 667 |
| Practical Assessment of Moderate Environments..... | 669 |
| Worker Relations.....                              | 669 |
| Expert Assessment .....                            | 669 |
| Where, When and What to Measure?.....              | 669 |
| Analysis.....                                      | 670 |
| Physical Measures .....                            | 670 |
| Prediction of Whole-Body Thermal Discomfort.....   | 671 |
| Local Thermal Discomfort .....                     | 673 |
| Subjective Responses .....                         | 674 |
| Concluding Remarks and Recommendations .....       | 674 |
| General Conclusion.....                            | 674 |
| Appendix.....                                      | 674 |
| References.....                                    | 676 |

## INTRODUCTION

This chapter presents the principles and practice of assessing environments for heat stress, cold stress and thermal comfort. Underpinning all of the assessment methods is that there are six important factors, all of which have to be taken into consideration when assessing human response to thermal conditions. These are the environmental factors: air temperature, radiant temperature, air velocity and humidity and the personal factors, clothing and activity. It is the interaction of all six factors that determine human response and not a subset or one alone. This can be demonstrated from heat transfer analysis between the human body and the environment (Parsons, 2014). Another principle is that a person responds to the local conditions to which they are exposed, so the aforementioned six factors have to be quantified in terms of the exposure of the person. The interaction of the four environmental factors can be considered to be the thermal exposure and of the six factors, the thermal stress. Thermal strain is the response of the body. A thermal index is a single number that represents the integrated thermal stress and can be used to predict thermal strain. Thermal strain can be physiological strain, measured by heart rate, body temperature, sweating, etc. and psychological strain including subjective response, such as ratings of discomfort or intolerance, or behavioural response such as changing clothing or moving away from the stress. The principle is that a strain is a 'movement away' from an optimal condition such as thermal comfort where there are preferred subjective responses as well as comfortable internal body temperatures and mean skin temperatures.

Applied ergonomics methods for assessing thermal environments include objective methods, subjective methods and methods using mathematical (usually computer) models. Objective methods include measuring the physiological response of people to the environment. Responses in terms of sweat rate, internal body temperature, skin temperatures and heart rate are useful measures of body strain (see Chapter 21). Performance measures at simulated or actual tasks can also be useful. Subjective measures are particularly helpful when assessing psychological factors such as thermal comfort and satisfaction. They can also be useful in quantifying the effects of moderate cold or moderate heat stress. Mathematical models have become popular in recent years, because, although often complex, they can be easily used in practical applications, employing computers. Some of the more sophisticated rational (or causal) models involve an analysis of the heat exchange between people and their environment and also include dynamic models of the human thermoregulatory system. Empirical models can provide useful mathematical equations which 'fit' data obtained from exposing human subjects to thermal conditions.

The aim of this chapter is to present the principles behind practical methods for assessing human response to hot, moderate and cold environments and to present a practical approach to assessing thermal environments with respect to human occupancy. It also presents mechanisms for determining safe surface temperatures so that products can be designed to avoid pain and burns on contact with the skin.

## THE PRINCIPLES

The following brief discussion provides the underlying principles behind assessing human response to thermal environments. For a fuller discussion and references, the reader is referred to a standard text such as Parsons (2014).

## RELEVANT MEASURES

It is now generally accepted that the six important factors that affect how people respond to thermal environments are air temperature, air velocity, radiant temperature, humidity and the clothing worn by, and the activity of, the human occupants of the environment. In any practical

assessment, instruments and methods for quantifying these factors may be used. Each of the instruments must meet specifications to ensure valid measurement. These specifications are provided in ISO 7726 (1998).

### **THERMOREGULATION**

People are homeotherms, that is, they react to thermal environmental stimuli in a manner which attempts to preserve their internal body ('core') temperature within an optimal range (around 37°C). If the body becomes too hot, vasodilation (blood vessel expansion) allows blood to flow to the skin surface (body 'shell'), providing greater heat loss. If vasodilation is an insufficient measure for maintenance of the internal body temperature, then sweating occurs, resulting in increased heat loss by evaporation. If the body becomes too cold, then vasoconstriction reduces blood flow to the skin surface and hence reduces heat loss to the environment. Shivering will increase metabolic heat production and can help maintain internal body temperature.

The physiological reaction of the body to thermal stress can have practical consequences. A rise or fall in internal body temperature can lead to confusion, collapse and even death. Vasoconstriction can lead to a reduction in skin temperature and complaints of cold discomfort and a drop in manual performance. Sweating can cause 'stickiness' and warmth, and 'mild heat' can provide a drop in arousal.

### **THERMAL INDICES**

A useful tool for describing, designing and assessing thermal environments is the thermal index. The principle is that factors that influence human response to thermal environments are integrated to provide a single index value. The aim is that the single value varies as human response varies and can be used to predict the effects of the environment. A thermal comfort index, for example, would provide a single number which is related to the thermal comfort of the occupants of an environment. It may be that two different thermal environments (i.e. with different combinations of various factors such as air temperature, air velocity, humidity and activity of the occupants) have the same thermal comfort index value. Although they are different environments, for an ideal index, identical index values would produce identical thermal comfort responses of the occupants. Hence, environments can be designed and compared using the comfort index.

A useful idea is that of the standard environment. Here the thermal index is the temperature of a standard environment that would provide the 'equivalent effect' on a subject as would the actual environment. Methods of determining equivalent effect have been developed. One of the first indices using this approach was the effective temperature (ET) index (Houghten and Yagloglou, 1923). The ET index was in effect the temperature of a standard environment (air temperature equal to radiant temperature, still air, 100% relative humidity for the activity and clothing of interest) which would provide the same sensation of warmth or cold felt by the human body as would the actual environment under consideration.

### **HEAT BALANCE**

The principle of heat balance has been used widely in methods for assessing human responses to hot, neutral and cold environments. If a body is to remain at a constant temperature, then the heat inputs to the body are balanced by the heat outputs. In fact, the body is in a dynamic equilibrium involving continuous interaction with the environment, responding in an attempt to maintain an optimum condition. Heat transfer can take place by conduction (*K*), convection (*C*), radiation (*R*) and evaporation (*E*). In the case of the human body, an additional heat input to the system is the

metabolic heat production ( $M$ ) due to the burning of food in oxygen by the body. Using the aforementioned, the following body heat equation can be proposed:

$$M \pm C \pm R \pm K - E = S. \quad (24.1)$$

If the net heat storage ( $S$ ) is zero, then the body can be said to be in heat balance and hence internal body temperature can be maintained. The analysis requires the values represented in Equation 24.1 to be calculated from a knowledge of the physical environment, clothing and activity.

### RATIONAL INDICES

Rational thermal indices use heat transfer equations (and sometimes mathematical representations of the human thermoregulatory system) to 'predict' human response to thermal environments. In hot environments, the heat balance equation (Equation 24.1) can be rearranged to provide the required evaporation rate ( $E_{req}$ ) for heat balance ( $S = 0$ ) to be achieved:

$$E_{req} = (M - W) + C + R. \quad (24.2)$$

( $K$  can often be ignored and  $W$  is the amount of metabolic energy that produces physical work.) Because sweating is the body's major method of control against heat stress,  $E_{req}$  provides a good heat stress index. A useful index related to this is to determine how wet the skin is; this is termed skin wettedness ( $w$ ), where

$$w = \frac{E}{E_{max}} = \frac{\text{Actual evaporation rate}}{\text{Maximum evaporation rate possible in that environment}}. \quad (24.3)$$

In cold environments, the clothing insulation required (IREQ) for heat balance can be a useful cold stress index based upon heat transfer equations.

Heat balance is not a sufficient condition for thermal comfort. In warm environments, sweating (or skin wettedness) must be within limits for thermal comfort, and in cold environments, skin temperature must be within limits for thermal comfort. Rational predictions of the body's physiological state can be used with empirical equations which relate skin temperature, sweat rate and skin wettedness to comfort.

### EMPIRICAL INDICES

Empirical thermal indices are based upon data collected from human participants who have been exposed to a range of environmental conditions. In hot environments, curves can be 'fitted' to sweat rates measured on individuals exposed to a range of hot conditions. There has been little research of this kind for cold conditions; however, a wind chill index was developed based upon the cooling of cylinders of water in outdoor conditions. Wind chill provides the 'trade-off' between air temperature and air velocity. Comfort indices have also been developed entirely empirically from subjective assessments over a range of environmental conditions.

### DIRECT INDICES

Direct indices are measurements taken on a simple instrument which responds to environmental components similar to those to which humans respond. For example, a wet, black globe with a thermometer placed at its centre will respond to air temperature, radiant temperature, air velocity and humidity. The temperature of the globe will therefore provide a simple thermal index



which, with experience of use, can provide a method of assessment of hot environments. Other instruments of this type include the temperature of a heated ellipse and the integrated value of an un aspirated wet bulb temperature, air temperature and black globe temperature Wet Bulb Globe Temperature (WBGT).

## MEASURING INSTRUMENTS

Air temperature is traditionally measured using a mercury-in-glass thermometer, although more recently, thermocouples and thermistors have become the preferred methods. An advantage of electronic instrumentation is that values can be continuously recorded and fed into computers for later analysis and if broken, then there is no exposure to toxic mercury. The dry bulb of a whirling hygrometer gives a value of air temperature. If there is a large radiant heat component in the environment, then it will be necessary to pass air over the sensor (e.g. by rapid whirling) or shield the air temperature transducer (e.g. using a wide-mouthed vacuum flask). Air humidity can be found from the wet and dry bulb of a whirling hygrometer. Other methods include capacitance devices and hair hygrometers (see ISO 7726, 1998).

Radiant temperature is usually quantified in the first analysis by measuring black globe (usually 150 mm diameter) temperature. Correcting the globe temperature for air temperature and air velocity allows a calculation of mean radiant temperature. If more detailed analysis is required, then instruments for measuring plane radiant temperatures in different directions should be used. Correction factors may be necessary to allow for the shape of the human body; however, the use of a globe thermometer provides a satisfactory initial measurement method.

Air velocity should be measured down to about  $0.1 \text{ m s}^{-1}$  in indoor environments. Generally cup or vein anemometers (i.e. masses rotated by moving air) will not measure such low air speeds. Suitable instruments are hot wire anemometers, where the cooling power of moving air over a hot wire, corrected for air temperature, provides air velocity, or Kata thermometers, where air movement cools a thermometer and cooling time is related to air velocity. Kata thermometers are however cumbersome and time consuming to use in practical applications.

In more recent years, integrating systems have been developed which detect all four environmental parameters and integrate measurements into thermal indices which predict, for example, thermal comfort. The instruments are usually simple to use and provide practical solutions for the non-expert. Another development has been the use of transducers connected to digital storage devices, to allow recording of environmental conditions over long periods of time. The devices usually allow easy interfacing with digital computers where sophisticated analysis can be performed. Technological advances allow new ways of achieving specifications and new opportunities for presenting measurements such as the use of data logging and local processing as well as the development of computer systems (including smartphone 'Apps'). It is important to remember the basic principles of physics when using a measuring instrument. Although apparently sophisticated and 'high spec', the position of the measurement device, measurement range, accuracy and response time will always be important and should not be ignored.

## SUBJECTIVE METHODS

Thermal comfort is usually defined as 'that condition of mind which expresses satisfaction with the thermal environment'. The reference to 'mind' emphasizes that this is a psychological phenomenon, hence the importance of subjective assessment methods. Subjective methods range from simple thermal sensation votes to more complex techniques where semantic or cognitive models of human perception of thermal environments can be determined. In a simple practical assessment of thermal environments, two types of scale are generally used: One type is concerned with thermal sensation and the other is concerned with acceptability (i.e. a value judgement). Specific questions regarding general satisfaction, draughts, dryness and open-ended questions asking for other comments provide

a useful brief subjective form, especially for measuring the acceptability and comfort of an environment. There are biases and errors which can occur in taking subjective measures, but it should not be forgotten that the best judges of their thermal comfort are the human occupants themselves. (See Chapter 4 on methods to collect participant responses and Chapter 23 on the environmental survey.)

### **BEHAVIOURAL AND ADAPTIVE METHODS**

In environmental assessment, the ergonomist should be aware that people are not simply passive receptors of the environment, but that they will respond to avoid discomfort and thermal strain. Any environmental assessment therefore should consider the 'adaptive opportunity' for people to alter their exposure to an environment. Can people move away from the environment, adjust clothing, adjust work rate, open windows, adjust heating controls, turn on fans and so on? Designing work to include adaptive opportunity can be part of an ergonomist's recommendations. Examples where adaptive opportunity may not exist or be low will be in hot environments where encapsulated protective clothing must be worn and the worker cannot easily leave the work (e.g. boiler stripping); in cold environments where activity level cannot be increased and in moderate environments where there is a strict dress code (e.g. uniform); in an emergency room where workers cannot leave the task or if windows cannot be opened or closed. Of particular concern will be populations who have limited adaptive opportunity due to their condition. These may include people with physical disabilities, old people, children and babies.

### **HUMAN PERFORMANCE**

Despite many studies into the effects of thermal environments on human performance, there is little specific information of practical value apart from cold conditions where manual dexterity and strength are clearly reduced with a significant drop in physical performance. The lack of consistency of findings in other areas is due to the contextual and individual factors involved which can have significant influence. Human performance can be considered in terms of physical (e.g. manual dexterity) and psychological (e.g. behavioural, cognitive) effects. If heat or cold stress is sufficiently severe that internal temperatures pass beyond limits at which major physiological effects occur (e.g. causing collapse or hallucinations), then clearly performance will be impaired. Within such limits, effects are influenced by factors such as motivation, level of proficiency at the task and individual differences.

The state of practical knowledge is such that it is not yet possible to predict reliably effects on manual or cognitive performance in hot environments or effects on cognitive performance in cold environments. Major effects do occur, however, on manual dexterity, cutaneous sensitivity and strength in cold environments. Hand skin temperature can be used to predict effects, and it is generally agreed that to maintain manual dexterity, the hands should be kept warm. If workers are sufficiently distracted to take time off the task, then this will clearly affect productivity. The distraction effects of thermal stress can also degrade performance, particularly in cold environments. Time off work due to environments exceeding heat or cold stress limits or time off task caused by distraction during work are reliable measures of loss in productivity. Predicting loss in capacity due to thermal conditions is less reliable because of individual differences, context and the many factors involved. That there are effects is clear, but general predictions without knowledge of these additional factors will not be reliable.

### **CLOTHING**

Clothing can be worn for protection against environmental hazards, for functional reasons (e.g. access to pockets) and for aesthetic reasons as well as for thermal insulation. In thermal terms a

microclimate is produced between the human body and the clothing surface. The applied ergonomist should ensure that the microclimate allows the body to achieve desirable physiological and psychophysical objectives. Physiological objectives may include the maintenance of heat balance for the body, and the preservation of skin temperatures and sweating at levels which allow for comfort. An interaction between thermal aspects and material type should also be considered (e.g. the effect of 'scratchy' materials being exacerbated by sweating).

The dry thermal insulation of clothing is greatly affected by how much air is trapped within clothing layers as well as within clothing. There are two common units which quantify dry clothing insulation; these are the Clo and the TOG. The Clo value is a clothing insulation value which is intended to be a 'relative unit', compared to a normal everyday costume necessary for thermal comfort in an indoor environment. For example, a 'typical' business suit (including underclothes, shirt) is often quoted as having a thermal insulation of 1.0 Clo. A nude person has 0 Clo. In terms of thermal insulation, 1 Clo is said to have a value of  $0.155 \text{ m}^2 \text{ }^\circ\text{C W}^{-1}$ . The TOG value is a unit of thermal resistance and is a property of the material. It can be measured on a heated flat plate, for example, in terms of heat transfer. It does not necessarily relate to the thermal insulation provided to a clothed person. For comparison, 1 TOG is equal to  $0.1 \text{ m}^2 \text{ }^\circ\text{C W}^{-1}$ . When clothing becomes wet, due to sweating or external conditions, then the clothing insulation is altered (usually greatly reduced). There are methods which estimate the thermal properties of wet clothing; however, these are crude and thermal insulation values for wet clothing are not well documented. Additional factors which can greatly affect the thermal insulation of clothing include pumping effects due to body movement, ventilation and wind penetration (see ISO 9920, 2009).

In application, it is not usually necessary to have a detailed description of clothing insulation properties. The important point is whether the clothing achieves its objectives. The objectives for clothing will be determined in an overall ergonomics systems analysis, involving a description of the objectives of the organization, task analysis, allocation of functions, job design and so on. This will include the objectives in terms of thermal insulation which will also involve the design of the clothing (to include pockets to keep hands warm, devices to keep workers cool in hot environments [e.g. ice jackets], etc.).

User tests and trials (described generally in Chapter 23) will provide important information about whether clothing meets its objectives. Objective measures such as sweat loss, skin temperatures and internal body temperatures or subjective measures of thermal sensation, comfort or stickiness can be of great value. Performance measures at actual or simulated tasks can also be used to evaluate whether clothing has achieved its thermal and other objectives.

## SAFE SURFACE TEMPERATURES

The applied ergonomist will be interested in the effect there will be on the body from physical contact between the human skin and surfaces in a workplace, for example, what sensation is caused by bare feet on a 'cold' floor or whether brief contact with a domestic product (e.g. a cooker, oven door, knobs or a kettle) will result in pain or burns. There are many factors involved in determining human response. These include the type and duration of contact, the material and condition of the surface and the condition of the human skin. However, a simple model based upon the heat transfer between two semi-infinite slabs of material in perfect thermal contact can provide a practical method of assessment.

The principal point is that there is a contact temperature which exists between the skin and the material surface, which is dependent upon the physical properties of the skin and the material, which will influence the effect on the person. For example, if one touched a metal slab at  $100^\circ\text{C}$ , then the contact temperature would be of the order of  $98^\circ\text{C}$  and the metal would be felt to be extremely hot. If, however, one touched a cork slab at  $100^\circ\text{C}$ , then the contact temperature would be around  $46^\circ\text{C}$ .

and the cork would feel much less hot than the metal. Contact temperature can therefore be used to predict effects on the body and can be calculated from the following equation:

$$T_{con} = \frac{(b_1 T_1 + b_2 T_2)}{(b_1 + b_2)}, \quad (24.4)$$

where

$T_1$  and  $T_2$  are the initial surface temperatures ( $^{\circ}\text{C}$ )

$T_{con}$  is the contact temperature ( $^{\circ}\text{C}$ )

$b_1$  and  $b_2$  are thermal penetration coefficients calculated from the following equation

$$b = (Kpc)^{1/2} \text{ J S}^{-1/2} \text{ m}^{-2} \text{ }^{\circ}\text{C}^{-1},$$

where

$K$  is the thermal conductivity

$p$  is density

$c$  is specific heat of the material

To calculate  $T_{con}$ , the values of  $b$  for human skin and for different materials are required. These are provided in Table 24.1.

McIntyre (1980) argues that if one assumes a skin temperature of  $34^{\circ}\text{C}$  and a simple reaction time for an individual of 0.25 s, then, from the work of Bull (1963), whose data suggest a partial burn at a skin temperature of about  $85^{\circ}\text{C}$  for 0.25 s contact, one can estimate temperatures which would produce partial burns. Some of these temperatures are provided in Table 24.1.

A practical approach to providing maximum surface temperatures for heated domestic equipment is provided in British Standard 4086 (BSI, 1966). The standard considers three categories of material type and three types of contact duration. A summary of the limits is provided in Table 24.2.

Because of the complexity of the problem and ethical considerations regarding experimentation involving pain and burns on human subjects, knowledge is incomplete in this area. In addition, specifications for safety of manufacturing products often involve other costs and benefits to be considered (see Chapter 35 in this book). The limits provided in BS 4086 (BSI, 1966) were reviewed and an additional document, BS PD 6504 (BSI, 1983), was produced which provides background medical information and data in terms of discomfort, pain and burns. There is, however, little information concerning the inter- and intra-participant variation in response, effects of skin condition and effects on different populations (e.g. the aged, children) and other

**TABLE 24.1**

**Thermal Penetration Coefficients ( $b$ )**

| Material   | $b$ ( $\text{J S}^{-1/2} \text{ m}^{-2} \text{ }^{\circ}\text{C}^{-1}$ ) | Momentary Contact Surface Temperature for Burn Threshold ( $^{\circ}\text{C}$ ) |
|------------|--------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Human skin | 1,000                                                                    | —                                                                               |
| Foam       | 30                                                                       | —                                                                               |
| Cork       | 140                                                                      | 450                                                                             |
| Wood       | 500                                                                      | 187                                                                             |
| Brick      | 1,000                                                                    | 136                                                                             |
| Glass      | 1,400                                                                    | 121                                                                             |
| Metals     | >10,000                                                                  | 90                                                                              |

Source: From McIntyre (1980).

**TABLE 24.2**  
**Maximum Surface Temperatures (°C) for Heated Domestic Equipment**

|                                                  | Handles (Kettles, Pans, etc.) | Knobs (Not Gripped) | Momentary Contact |
|--------------------------------------------------|-------------------------------|---------------------|-------------------|
| Metals                                           | 55                            | 60                  | 105               |
| Vitreous enamelled steel<br>and similar surfaces | 65                            | 70                  | 120               |
| Plastics, rubber or wood                         | 75                            | 85                  | 125               |

Source: BSI, *Recommendations for Maximum Surface Temperatures of Heated Domestic Equipment*, British Standards Institution, London, U.K., 1966.

variables important for practical application. Further work in this area has led to European and International Standards (EN 563, ISO 13732).

ISO 13732 is in three parts: part 1 for skin contact with hot surfaces, part 2 for contact with moderate surfaces and part 3 for cold surfaces. For a full discussion of human skin reaction to contact with hot, moderate and cold surfaces, see Parsons (2014). This is an important area for product design where a wide range of products have specified temperature limits to avoid pain and injury to users.

**INTERNATIONAL STANDARDS**

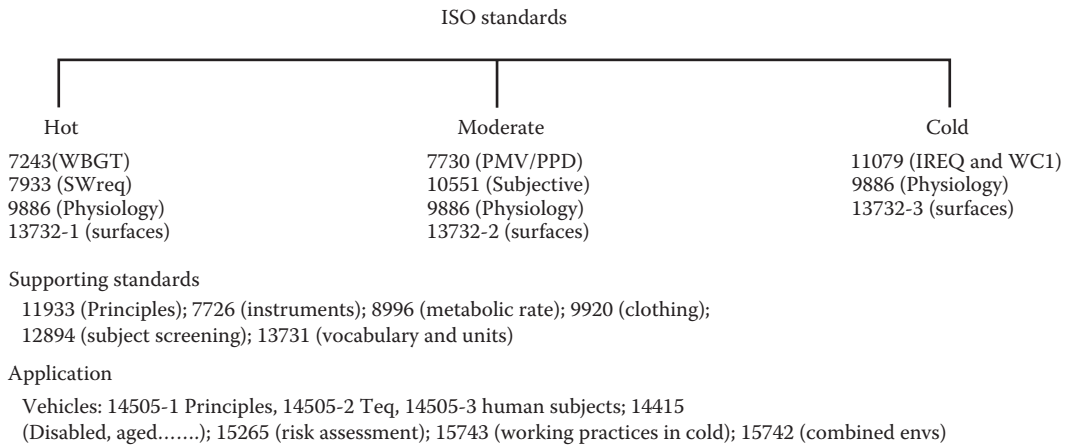
There have been many national and international standards concerned with thermal comfort, heat stress and cold stress. Thermal comfort standards can define conditions for thermal comfort and indicate the likely degree of discomfort of occupants of thermal environments. Standards for heat stress and cold stress attempt to specify conditions that will preserve health and often comfort and performance. Standards can also provide guidance on environmental design and control, they standardize methods to allow comparison and they contribute to assessment and evaluation.

Influential institutions throughout the world that produce standards or guidelines include the *American Society of Heating, Refrigerating and Air Conditioning Engineers* (ASHRAE) in the United States, the *Chartered Institute of Building Services Engineers* (CIBSE) in the United Kingdom, national standards bodies, the ISO and also European Standardization under CEN, the *World Health Organization* (WHO), the *American Conference of Governmental Industrial Hygienists* (ACGIH), and the *International Labour Organization* (ILO). Many standards have defined limits or at least methods in terms of rational or empirical thermal indices or physiological condition (e.g. body ‘core’ temperature). There are also standards that provide techniques and methods (e.g. physiological and subjective assessment methods).

**ISO STANDARDS**

For the assessment of hot environments, a simple method based on the WBGT index is provided in ISO 7243. If the WBGT reference value is exceeded, a more detailed analysis can be made (ISO 7933) involving calculation, from the heat balance equation, of predicted rectal temperature and sweating required in a hot environment. If the responses of individuals, or of specific groups, are required (e.g. in extremely hot environments), then physiological strain should be measured. Methods of measuring mean skin temperature, heart rate, internal body (‘core’) temperature and mass loss are all described in ISO 9886 (Figure 24.1).

ISO 7730 provides an analytical method for assessing moderate environments and is based on the predicted mean vote and predicted percentage of dissatisfied (PMV/PPD) index and on criteria for local thermal discomfort, in particular draughts. If the responses of individuals or specific groups are required, then subjective measures should be used (ISO 10551).



**FIGURE 24.1** ISO standards for assessing thermal environments.

ISO 11079 provides an analytical method for assessing cold environments involving calculation of the clothing insulation required (IREQ) from a heat balance equation. This can be used as a thermal index or as a guide for selecting clothing.

Supporting standards include an introductory standard (ISO 11399) and standards for estimating the thermal properties of clothing (ISO 9920) and metabolic heat production (ISO 8996) and a standard for definitions, symbols and units (ISO 13731). Other standards consider instruments and measurement methods (ISO 7726) and medical supervision of individuals exposed to hot or cold environments (ISO 12894). ISO work on contact with solid surfaces is divided into hot, moderate and cold surfaces. Ergonomics/human factors (E/HF) data to establish temperature limit values for hot touchable surfaces are provided in European standard EN 563 (1994). This has been extended to the three-part standard ISO13732 (part 1, hot surfaces; part 2, moderate surfaces and part 3, cold surfaces). Newly developed standards include a four-part standard for the assessment of vehicle thermal environments (ISO 14405), a standard for people with special requirements (e.g. disabled, aged, ISO TS 14515) and a standard for risk assessment (ISO 15265). Working practices for cold environments are presented in ISO 15743.

ISO 11399 (1995) presents a description of principles and methods of application of the series of ISO standards and should be consulted for an initial overview. The ISO working system showing how the collection of standards can be used in practice is presented in Figure 24.2. The objectives and status of the standards are shown in Table 24.3.

## THERMAL MODELS

Practical models which simulate how people respond to hot, moderate or cold environments can be used to assess and design thermal environments. They can also be integrated into larger computer-based expert and knowledge-based systems for use by the ergonomics practitioner.

Examples of models which simulate the human response to thermal environments are provided by Haslam and Parsons (1987), Fiala et al. (2012) and Parsons (2014). Models of the human thermoregulatory system controlling a passive body (e.g. made up of cylinders and a sphere with thermal properties similar to those of the human body) can be used to predict changes in temperature within and over different parts of a clothed body. These models can then simulate how persons could respond in terms of heat stress or cold stress in outdoor environments or in terms of thermal comfort indoors. Investigations of the nature of expertise used in assessing the human response to thermal environments, coupled with the requirements of ergonomics practitioners and simulations using such models as are described here, can



| Type of thermal environment      |                                                                             |                                                                              |                                                         |
|----------------------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------------------|---------------------------------------------------------|
|                                  | Hot                                                                         | Moderate                                                                     | Cold                                                    |
| Principles and application       |                                                                             |                                                                              |                                                         |
| Evaluation of comfort and stress | Wet bulb globe temperature index, WBGT Required Sweat Rate (SWreq)          | predicted mean vote, PMV and predicted percentage dissatisfied, PPD, indices | Wind chill index, WCI Required clothing insulation IREQ |
| Strain evaluation                |                                                                             |                                                                              |                                                         |
| Physiological                    | 'Core' and skin temperature, heart rate, mass lost sweating and respiration |                                                                              |                                                         |
| Psychological                    | Subjective assessment methods                                               |                                                                              |                                                         |

**FIGURE 24.2** Assessment of thermal environments using the series of ISO standards.

provide the input to expert systems. An example of the structure of such a system as described by Smith and Parsons (1987) is provided in Figure 24.3. Such systems have become valuable tools in integrating the principles and knowledge involved in the assessment of human response to thermal environments for practical application. It is surprising that such systems are not more widely used, especially with the utility of the Internet and the World Wide Web. They remain a source of great potential for the assessment of thermal environments and the general application of environmental ergonomics.

## THE PRACTICE

The principles mentioned earlier can provide a practical methodology which can be used to assess thermal environments with respect to effects on their human occupants. Almost by definition, practical assessments will have factors specific to particular applications and one universal method is therefore difficult to provide.

## PRACTICAL ASSESSMENT OF HOT ENVIRONMENTS

The assessment of hot environments is particularly important as danger to health can occur rapidly. For environments where experience has been gained in monitoring workers, working practices can be developed and conditions monitored using simple thermal indices (e.g. WBGT index).

The WBGT index is used in ISO 7243 as a simple method for assessing hot environments. For conditions inside buildings and outside buildings without solar load,

$$\text{WBGT} = 0.7t_{nw} + 0.3t_g, \quad (24.5)$$

and outside buildings with solar load,

$$\text{WBGT} = 0.7t_{nw} + 0.2t_g + 0.1t_a, \quad (24.6)$$

where

$t_{nw}$  is the natural wet bulb temperature (i.e. not 'whirled')

$t_g$  is 150 mm diameter black globe temperature

$t_a$  is the air temperature

**TABLE 24.3****Published ISO Standards for the Assessment of Human Response to the Thermal Environment**

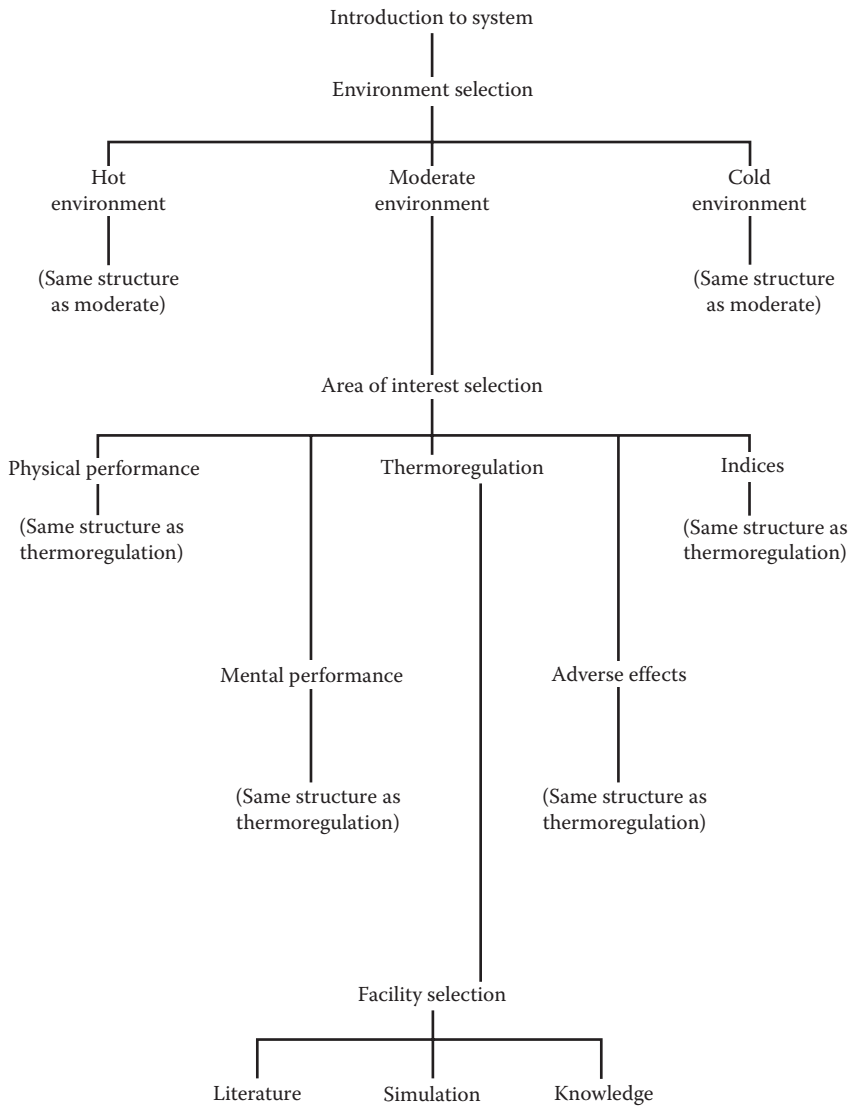

---

|                              |                                                                                                                                                                                                                     |
|------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO 7243 (1989) (ED 2)       | Hot environments – Estimation of the heat stress on working man, based on the WBGT index                                                                                                                            |
| ISO 7266 (1998) (ED 2)       | Ergonomics of the thermal environment – Instruments for measuring physical quantities                                                                                                                               |
| ISO 7730 (2005) (ED 3)       | Ergonomics of the thermal environment – Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria                              |
| ISO 7933 (2004) (ED 2)       | Ergonomics of the thermal environment – Analytical determination and interpretation of heat stress using calculation of the predicted heat strain                                                                   |
| ISO 8996 (2004) (ED 2)       | Ergonomics of the thermal environment – Determination of metabolic rate                                                                                                                                             |
| ISO 9886 (2004) (ED 2)       | Evaluation of thermal strain by physiological measurements                                                                                                                                                          |
| ISO 9920 (2007) (ED 2)       | Estimation of thermal insulation and water vapour resistance of a clothing ensemble                                                                                                                                 |
| ISO 10551 (1995) (ED 1)      | Ergonomics of the thermal environment – Assessment of the influence of the thermal environment using subjective judgement scales                                                                                    |
| ISO 11079 (2007) (ED 1)      | Ergonomics of the thermal environment – Determination and interpretation of cold stress when using required clothing insulation (IREQ) and local cooling effects                                                    |
| ISO 11399 (1995) (ED 1)      | Ergonomics of the thermal environment – Principles and application of relevant international standards                                                                                                              |
| ISO 12894 (2001) (ED 1)      | Ergonomics of the thermal environment – Medical supervision of individuals exposed to extreme hot or cold environments                                                                                              |
| ISO 13731 (2001) (ED 1)      | Ergonomics of the thermal environment – Vocabulary and symbols                                                                                                                                                      |
| ISO 13732-1 (2001) (ED 1)    | Ergonomics of the thermal environment – Methods for the assessment of human responses to contact with surfaces – Part 1: Hot surfaces                                                                               |
| ISO/TS 13732-2 (2001) (ED 1) | Ergonomics of the thermal environment – Methods for the assessment of human responses to contact with surfaces – Part 2: Human contact with surfaces at moderate temperature                                        |
| ISO 13732-3 (2005) (ED1)     | Ergonomics of the thermal environment – Methods for the assessment of human responses to contact with surfaces – Part 3: Cold surfaces                                                                              |
| ISO/TS 14505-1 (2007) (ED 1) | Ergonomics of the thermal environment – Evaluation of the thermal environment in vehicles – Part 1: Principles and methods for assessment of thermal stress                                                         |
| ISO 14505-2 (2006) (ED 1)    | Ergonomics of the thermal environment – Evaluation of the thermal environment in vehicles – Part 2: Determination of equivalent temperature (see also ISO 14505-2, 2006/Cor 1:2007: [ED 1] Technical Corrigendum 1) |
| ISO 14505-3 (2006) (ED 1)    | Ergonomics of the thermal environment – Evaluation of the thermal environment in vehicles – Part 3: Evaluation of thermal comfort using human subjects                                                              |
| ISO 15265 (2004) (ED 1)      | Ergonomics of the thermal environment – Risk assessment strategy for the prevention of stress or discomfort in thermal working conditions                                                                           |
| ISO 15743 (2008) (ED 1)      | Ergonomics of the thermal environment – Cold work places – Risk assessment and management                                                                                                                           |

---

Acclimatization programmes, before workers begin work, are useful. It is particularly important that workers do not become unacceptably dehydrated (e.g. >4% of body weight lost in sweat) or have an unacceptably elevated internal body temperature (e.g. >38.0°C–38.5°C). A more detailed analysis can be provided by using rational assessments of the environment (ISO 7933). Allowable exposure times based on such factors as predicted elevated internal body temperature or dehydration can be provided. It is important to remember however that there are individual differences in workers and that knowledge of heat transfer for the human body is incomplete. Experience is therefore required in the use of rational models.

If individuals are exposed to extremely hot environments, then individual physiological measures of heart rate, internal body temperature and sweat loss should be taken and each worker observed closely.



**FIGURE 24.3** System tree diagram.

## PRACTICAL ASSESSMENT OF COLD ENVIRONMENTS

Similar general guidelines apply to cold environments as were described for hot environments. A simple index such as the wind chill index can be used when experience has been gained with its use, where

$$\text{WCI} = (10\sqrt{v} + 10.45 - v) (33 - t_a), \quad (24.7)$$

where

$v$  is the air velocity ( $\text{m s}^{-1}$ )

$t_a$  is the air temperature ( $^{\circ}\text{C}$ )

The effects associated with different values of the wind chill index are

| WCI  | Effect                |
|------|-----------------------|
| 200  | Pleasant              |
| 400  | Cool                  |
| 1000 | Cold                  |
| 1200 | Bitterly cold         |
| 1400 | Exposed flesh freezes |
| 2500 | Intolerable           |

When the WCI value is calculated, it is often useful to calculate the  $t_a$  value which would provide the same wind chill in 'calm' air ( $v = 1.8 \text{ m s}^{-1}$ ). This is called the chilling temperature and is given by

$$t_{ch} = \frac{33 - \text{WCI}}{22} ^\circ\text{C}. \quad (24.8)$$

This index is still widely used across the world but has also been replaced in many countries by the IREQ index for whole-body cooling and the wind chill temperature for local cooling of the skin (ISO 11079). The IREQ index is a rational index, calculated from the heat balance equation. The wind chill temperature ( $t_{WC}$ ) is given by

$$t_{WC} = 13.12 + 0.6215t_a - 11.37v_{10}^{0.16} + 0.3965t_a v_{10}^{0.16}. \quad (24.9)$$

The wind chill temperature ( $t_{WC}$ ) is the ambient temperature which at a wind speed of  $4.2 \text{ km h}^{-1}$  produces the same cooling power (sensation) as the actual environmental conditions. It should be noted that  $v_{10}$  is a meteorological measure and is the air velocity at the top of a 10 m pole. Ergonomists usually measure the air velocity at the position of the person, at ground level. Air velocity at 10 m is estimated by multiplying air velocity at ground level by 1.5.  $t_{WC}$  values between  $-10^\circ\text{C}$  and  $24^\circ\text{C}$  are rated as 'uncomfortably cold'; between  $-12^\circ\text{C}$  and  $-34^\circ\text{C}$ , 'very cold' with a risk of skin freezing;  $-35^\circ\text{C}$  to  $-59^\circ\text{C}$ , 'bitterly cold' where exposed skin may freeze in 10 min; and  $-60^\circ\text{C}$  and colder, 'extremely cold' where exposed skin may freeze within 2 min.

Clothing is important, and a compromise must be reached between thermal insulation and clothing design to reduce effects on worker performance and safety. Of particular interest is the temperature of the body's extremities (hands and feet). Prolonged exposure may lead to thermal injury, but severe discomfort, distraction and loss of manual dexterity are the most commonly occurring effects. Although there is some debate, 'back of hand' temperatures above  $20^\circ\text{C}$ – $25^\circ\text{C}$  should maintain some comfort and performance. Hand temperatures of  $<10^\circ\text{C}$ – $15^\circ\text{C}$  are usually unsatisfactory, although they should not produce injury. Performance effects will depend upon duration of exposure. Cold can produce severe discomfort and this has behavioural and distractive effects. Distraction may reduce manual and cognitive workload capacity.

Rational indices can be used to predict the required clothing insulation (IREQ) for heat balance and thermal comfort and also allowable exposure times based upon a drop in body 'core' temperature. Physiological measures of heart rate, mean skin temperature and body core temperature should be used if assessing individuals in cold environments. Any drop in body core temperature is

unsatisfactory, but 36°C is a lower working limit. A core temperature of below 35°C is defined as hypothermia. Medical screening of subjects should take place before exposure to either hot or cold environments (ISO 12894).

## **PRACTICAL ASSESSMENT OF MODERATE ENVIRONMENTS**

A common request to the environmental ergonomist is to assess an indoor climate such as found in an office. The practical method used in an actual case is outlined here, although some adjustment to the results has been made to illustrate points.

The workers in a large office were complaining that their thermal environment was unacceptable. The ergonomist was asked to assess the environment, quantify the problem and make recommendations for improvement, if necessary. One day was allowed for the assessment and a total of 4 days for the whole project, including both analysis and final report, a not unusual time restriction.

### **WORKER RELATIONS**

Complaints about working environments can be stimulated by other work-related problems, and it is important for the ergonomist to gain an impression of the physical, social and organizational environment in general. In addition, it is useful to have the co-operation and understanding of management and workers. The worker representative was therefore contacted and the ergonomist introduced. It was explained that the ergonomist was attempting to improve the thermal environment conditions. The physical and subjective measures which were to be taken were also demonstrated to the workers' representative who then passed on the information to the office occupants.

### **EXPERT ASSESSMENT**

An initial expert assessment was made using a simple checklist similar to that shown in Chapter 23. As well as general first impression, of particular importance was the identification of adaptive opportunities in the workplace. These can be divided into opportunities to keep warm or opportunities to cool down. Clothing adjustment was possible in the environment and it was possible to move around and change posture. It was estimated that adaptive opportunities to keep warm and to cool down could be regarded as high as long as the workers were aware of the opportunities and encouraged to take advantage of them. This observation was important in the final recommendation.

### **WHERE, WHEN AND WHAT TO MEASURE?**

The question of where and when to measure is a question of statistical sampling. The thermal environmental conditions will vary throughout a space and also with time (during the day, night and seasonal variations). The more measuring points in the room and the more measuring times, in general, the more accurately the environment can be quantified. This then is a question of resources. Only one day was allowed for measurement, so a plan of the office was obtained and individual workplaces identified. Measurements should be taken at the positions of the workers. Ankle, chest and head heights were chosen as measuring points at each workplace. Ten workplaces were chosen as the sample, 'evenly spread' throughout the office. Measurements were taken over a 3 h period under what had been established as 'typical' conditions throughout the morning, a time when complaints had been received. The ventilation systems were identified and set to normal working. Outside weather conditions were noted.

A 150 mm diameter globe thermometer was placed at each workplace (only two were available, so they had to be moved around) for at least 20 min before readings were taken. Using a hot wire anemometer, air velocity and air temperature were measured at ankle, chest and head height of the worker. A whirling hygrometer was used at chest height to measure wet and dry bulb temperatures (dry bulb was used as a cross-check for air temperature with the air temperature sensor on the hot wire anemometer). The workers' clothing and activity were noted, and movements throughout the room were also noted.

Subjective assessment forms (see the Appendix) were handed to each worker and collected centrally (i.e. the working position was noted, but a degree of anonymity was maintained). The subjective forms allowed some information to be collected regarding time variations (i.e. outside the survey time) and general satisfaction.

ANALYSIS

Physical Measures

Analysis of physical measurements takes place in two parts. The first part is to obtain for each measurement point air temperature, mean radiant temperature, relative humidity and air velocity from the instrument measures and also to determine metabolic heat production and clothing insulation values. The second part is to predict the degree of discomfort. The subjective measures are analysed separately and complement the physical measures.

The air temperature and air velocity were measured directly using the hotwire anemometer. The mean radiant temperature  $t_r$  is obtained from the globe temperature  $t_g$  corrected for air temperature ( $t_a$  and air velocity  $[v]$ ). If the mean radiant temperature is within a few degrees of room temperature, then McIntyre (1980) suggests that

$$t_r = t_g + 2.44\sqrt{v} (t_g - t_a), \tag{24.10}$$

where temperatures are in °C and air velocity in m s<sup>-1</sup>. Relative humidity is calculated from the dry bulb (air temperature) and aspirated (whirled) wet bulb of the whirling hygrometer. Table 24.4 provides typical values. Table 24.5 provides metabolic heat production values for typical activities and

**TABLE 24.4**  
**Relative Humidity (%) from Dry Bulb and Aspirated Wet Bulb Temperature**

| Dry Bulb Temperature (°C) | Aspirated Wet Bulb Temperature (°C) |     |     |     |     |     |     |     |     |     |
|---------------------------|-------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                           | 12                                  | 14  | 16  | 18  | 20  | 22  | 24  | 26  | 28  | 30  |
| 12                        | 100                                 |     |     |     |     |     |     |     |     |     |
| 14                        | 79                                  | 100 |     |     |     |     |     |     |     |     |
| 16                        | 62                                  | 81  | 100 |     |     |     |     |     |     |     |
| 18                        | 49                                  | 64  | 82  | 100 |     |     |     |     |     |     |
| 20                        | 37                                  | 51  | 66  | 83  | 100 |     |     |     |     |     |
| 22                        | 28                                  | 40  | 54  | 68  | 83  | 100 |     |     |     |     |
| 24                        | 20                                  | 31  | 43  | 56  | 69  | 84  | 100 |     |     |     |
| 26                        | 14                                  | 24  | 34  | 45  | 58  | 71  | 85  | 100 |     |     |
| 28                        | 9                                   | 18  | 27  | 37  | 48  | 59  | 72  | 85  | 100 |     |
| 30                        | 5                                   | 12  | 21  | 30  | 39  | 50  | 61  | 73  | 86  | 100 |



**TABLE 24.5**  
**Estimates of Typical Metabolic Heat Production Values**

| Activity                 | Metabolic Heat Production (W m <sup>-2</sup> ) |
|--------------------------|------------------------------------------------|
| Seated, at rest          | 58                                             |
| Standing, relaxed        | 70                                             |
| Standing, light arm work | 100                                            |
| VDU operation            | 70                                             |
| Driving                  | 70–100                                         |

**TABLE 24.6**  
**Estimates of Typical Clothing Insulation Values**

| Type of Clothing                                                                                        | Clothing Insulation (Clo) |
|---------------------------------------------------------------------------------------------------------|---------------------------|
| None                                                                                                    | 0                         |
| Light summer clothing (briefs, shorts, short-sleeved shirt, light socks, light shoes)                   | 0.3                       |
| Light work clothing (light underwear, cotton long-sleeved workshirt, light long trousers, socks, shoes) | 0.65                      |
| Light business suit (including underclothing)                                                           | 1.0                       |
| Heavy business suit (including underclothing)                                                           | 1.5                       |

1 Clo = 0.155 m<sup>2</sup> °C W<sup>-1</sup>.

Table 24.6 provides clothing insulation values for typical clothing. Useful information is provided by presenting all physical data in a table, or on the office plan in the final report.

**Prediction of Whole-Body Thermal Discomfort**

One of the most useful thermal comfort indexes is the PMV of Fanger (1970) which is used in ISO 7730. Air temperature, mean radiant temperature, air velocity, humidity, clothing and activity values can be integrated to predict the mean thermal sensation vote of a large group of people on a seven-point thermal sensation scale (as used on the subjective assessment form in the Appendix). The values range from PMV = 3 (hot) through PMV = 0 (neutral) to PMV = -3 (cold). PMV = 0 provides comfort conditions. From the PMV value, a PPD value can be calculated. This is related to the percentage of people likely to complain about the thermal conditions. Values of PMV and PPD are presented in Tables 24.7 and 24.8 for typical environmental conditions.

The thermal discomfort results for all 10 workplaces will not be presented here; however, the PMV and PPD values for each workplace were calculated and labelled on a copy of the plan of the office, for the final report. This showed the predicted whole-body thermal sensation (comfort) pattern over the office and those areas of likely complaint.

The following is an example of calculations for one workplace; the physical measurements were as follows:  $t_a = 18^{\circ}\text{C}$ ,  $t_r = 18^{\circ}\text{C}$ ,  $v = 0.15\text{ m s}^{-1}$ , relative humidity = 50%, clothing insulation = 0.65 Clo, and metabolic rate =  $70\text{ W m}^{-2}$ .

Using Tables 24.7 and 24.8,

$$\text{PMV} = -1.7 \text{ and } \text{PPD} = 62\%.$$

**TABLE 24.7**  
**PMV Values for a Range of Air Temperature, Clothing and Activity Values**

| Clothing | Activity (W m <sup>-2</sup> ) | Air Temperature (°C) |      |      |      |      |      |     |
|----------|-------------------------------|----------------------|------|------|------|------|------|-----|
|          |                               | 16                   | 18   | 20   | 22   | 24   | 26   | 28  |
| 0.5      | 58                            | —                    | —    | -2.5 | -1.8 | -1.0 | -0.2 | 0.6 |
| 0.65     | 58                            | —                    | -2.7 | -2.0 | -1.3 | -0.6 | 0.0  | 0.8 |
| 1.0      | 58                            | -2.1                 | -1.6 | -1.1 | -0.5 | 0.0  | 0.6  | 1.2 |
| 1.5      | 58                            | -1.1                 | -0.7 | -0.3 | 0.2  | 0.6  | 1.1  | 1.5 |
| 0.2      | 70                            | —                    | —    | -2.8 | -2.0 | -1.2 | -0.4 | 0.4 |
| 0.5      | 70                            | -2.8                 | -2.2 | -1.6 | -1.0 | -0.4 | 0.3  | 0.9 |
| 0.65     | 70                            | -2.2                 | -1.7 | -1.2 | -0.6 | 0.0  | 0.5  | 1.0 |
| 1.0      | 70                            | -1.3                 | -0.9 | -0.5 | 0.0  | 0.4  | 0.9  | 1.3 |
| 1.5      | 70                            | -0.5                 | -0.2 | 0.2  | 0.5  | 0.9  | 1.2  | 1.6 |
| 0.5      | 100                           | -1.2                 | -0.8 | -0.4 | 0.0  | 0.5  | 0.9  | 1.4 |
| 0.65     | 100                           | -0.9                 | -0.5 | -0.1 | 0.3  | 0.6  | 1.0  | 1.4 |
| 1.0      | 100                           | -0.3                 | 0.0  | 0.3  | 0.6  | 1.0  | 1.3  | 1.6 |
| 1.5      | 100                           | 0.3                  | 0.5  | 0.7  | 1.0  | 1.3  | 1.5  | 1.8 |

*Notes:* Assume mean radiant temperature = air temperature; air velocity = 0.15 m s<sup>-1</sup>; relative humidity = 50% (3, hot; 2, warm; 1, slightly warm; 0, neutral; -1, slightly cool; -2, cool; -3, cold).

**TABLE 24.8**  
**Interpretation of PMV Values in Terms of Thermal Sensation and PPD**

| Sensation | Cold | Cool | Slightly Cool | Neutral | Slightly Warm | Warm | Hot |
|-----------|------|------|---------------|---------|---------------|------|-----|
| PMV       | -3   | -2   | -1            | 0       | 1             | 2    | 3   |
| PPD (%)   | —    | 75   | 25            | 5       | 25            | 75   | —   |

Therefore, the prediction is that, on average, a person will be between slightly cool and cool at this position. Also it can be seen that for all other conditions remaining the same, an increase in air temperature from 18°C to 24°C (or an increase in clothing insulation to around 1.6 Clo) will provide a PMV value of 0 required for comfort. This could be a recommendation, or a recommendation could be made in terms of increased activity level or a combination, etc.

In terms of using adaptive opportunity, it had been assessed (earlier) that this was high, so a range of thermal conditions will still allow the workers to achieve comfort. A method of taking account of that ( $I_{EQUIV}$ ) using the PMV index can give an indication of the range of temperatures that will provide thermal comfort as long as the workers take advantage of the opportunities.

Two equations are used, one for adaptation to heat and the other to cold.

When adapting to heat and adaptation is equivalent to reducing clothing insulation,

$$I_{EQUIV} = I_{START} - (I_{ADJ} \times I_{START}). \quad (24.11)$$

When adapting to cold and adaptation is equivalent to increasing clothing insulation,

$$I_{EQUIV} = I_{START} + (I_{ADJ} \times I_{START}). \quad (24.12)$$

Where

$I_{EQUIV}$  is the equivalent clothing index and is defined as the clothing insulation that would give equivalent thermal comfort to people with no adaptation as the thermal comfort of people who adapt to their thermal conditions

$I_{START}$  is the clothing insulation that people 'start' with and hence the value that is adjusted by adaptation. In the case of this example, it is 0.65 Clo

$I_{ADJ}$  is the adjustment made to the clothing due to the adaptive opportunity provided by the environmental design. An estimate is made of how much clothing adjustment would be equivalent to the total of all adaptive behaviours. Although specific calculations can be complex, in practical applications, complexity is not necessary and a simple estimate of  $I_{ADJ}$  based upon adaptive opportunity is 0, minimum; 0.25, low; 0.5, medium; 0.75, high; and 1.0, maximum. Clothing values are in Clo

For the current office, it can be calculated (see Table 24.7) that for comfort ( $PMV = 0$ ), air temperature should be kept at 24°C, but that assumes no adaptation by the workers. Adaptive opportunity has been estimated from the expert assessment, as high ( $ADJ = 0.75$ ) for when conditions become too warm or too cool.

For warm conditions from Equation 24.11,

$$I_{EQUIV} = 0.65 - (0.75 \times 0.65) = 0.1625 \text{ Clo.}$$

For cold conditions from Equation 24.12,

$$I_{EQUIV} = 0.65 + (0.75 \times 0.65) = 1.1375 \text{ Clo.}$$

For comfort ( $PMV = 0$ ) for identical conditions but for people wearing 0.1625 Clo (i.e. just below 0.2 Clo), an air temperature of 27°C will be sufficient for cooling in hot conditions, and for 1.1375 Clo (just above 1.0 Clo), an air temperature of 21°C will be sufficient for heating in cool conditions. The difference between cooling only to 27°C instead of a more prescriptive 24°C in hot conditions will be significant in terms of energy demand as will heating to only 21°C instead of to 24°C in cold conditions. It is important however to impress upon the manager of the workspace to ensure that the workers are allowed to adapt to their thermal conditions and that they are trained to do so.

### Local Thermal Discomfort

As well as overall or whole-body thermal sensation, thermal conditions can produce effects on local areas of the body. For example, cold air moving around the workers' ankles may cause a draught. The most common forms of local discomfort are caused by cooling due to air movement, heat losses due to asymmetric radiation (e.g. a radiant draught caused by workers sitting next to cold walls or windows) and thermal gradients. There is some debate about conditions which produce discomfort, but cool air movements (especially, if fluctuating) should be avoided above 0.15 m s<sup>-1</sup> and particularly for exposed skin areas and if the subject is already cool. Radiant asymmetry should not exceed 10°C (less in the case of heated ceilings) and vertical temperature gradients should not be >3°C. General observation of the workplaces, air velocity measures at the three heights (ankle, chest and head) and mean radiant temperatures will provide an indication of possible local thermal discomfort.

Dryness is probably related to air velocity, humidity and air and radiant temperatures, and is usually due to the evaporation of fluids from the eyes, nose and mouth which can lead to various problems, for example, with contact lenses. Local discomfort and other factors such as dryness and overall satisfaction should also be examined using subjective methods.

### **Subjective Responses**

Analysis of subjective responses involves determining the average of, and variation in, response. The responses of how workers felt at the time of measurement can be compared with predicted responses. In general, subjects in the office example used gave a wider range on the scale than the predicted measures. The subjective measures were also presented on a plan of the office in the final report. On average, workers were between slightly cool and cool, with some participants cold and some neutral. Draughts were reported in some areas. Most workers wished to be warmer. Responses regarding general sensation at work were similar to responses made about the conditions when they were measured. Most people were generally dissatisfied with the thermal environment.

## **CONCLUDING REMARKS AND RECOMMENDATIONS**

The aforementioned measurement and analysis allowed recommendations to be made in a final report which were related to the original objectives. An average increase in air temperature was recommended with some specific recommendations about draughts for particular workstations. It was also noted that the high level of dissatisfaction indicated may be due to general work or workplace dissatisfaction and not simply related to thermal conditions.

### **GENERAL CONCLUSION**

The aforementioned example provides a simple but effective practical method for assessing environments. Future trends will include improved measurement technology (instruments, subjective responses, behavioural responses) using local storage or online remote methods. The ergonomist is in a unique position to take a holistic view of design and assessment and not only consider the thermal environment in isolation but as an integral part of a complex system.

## **APPENDIX**

The following subjective form was used in a moderate office environment where workers had been complaining about general working conditions. Various details about workers' characteristics and location were collected separately. The form was handed to workers for completion at their workplace. Question 1 determines the workers' sensation vote on the ISO scale. Note that this can be compared directly with the measured PMV. Question 2 provides an evaluation judgement. For example, question 1 determines subject's sensation (e.g. warm). Question 2 compares this sensation with how the subject would like to be. Questions 3 and 4 provide information about how workers generally find their thermal environment. This is useful where it is not practical to survey the environment for long durations. Questions 5 and 6 are catch-all questions about workers' satisfaction and any other comments. Answers to these questions will provide information about whether more detailed investigation is required. Answers will also indicate factors which are obvious to the workers but not obvious to the investigator.

**Form 1**

Please answer the following questions concerned with YOUR THERMAL COMFORT.

1. Indicate on the scale below how you feel NOW.

|               |       |
|---------------|-------|
| Hot           | _____ |
| Warm          | _____ |
| Slightly warm | _____ |
| Neutral       | _____ |
| Slightly cool | _____ |
| Cool          | _____ |
| Cold          | _____ |

2. Please indicate how you would like to be NOW.

|        |           |        |
|--------|-----------|--------|
| Warmer | No change | Cooler |
|--------|-----------|--------|

3. Please indicate how you GENERALLY feel at work:

|               |       |
|---------------|-------|
| Hot           | _____ |
| Warm          | _____ |
| Slightly warm | _____ |
| Neutral       | _____ |
| Slightly cool | _____ |
| Cool          | _____ |
| Cold          | _____ |

4. Please indicate how you would GENERALLY like to be at work:

|        |           |        |
|--------|-----------|--------|
| Warmer | No change | Cooler |
|--------|-----------|--------|

5. Are you generally satisfied with your thermal environment at work?

|     |    |
|-----|----|
| Yes | No |
|-----|----|

6. Please give any additional information or comments which you think are relevant to the assessment of your thermal environment at work (e.g. draughts, dryness, clothing, suggested improvements, *etc.*)

Thank you

## REFERENCES

- BSI (1966) *Recommendations for Maximum Surface Temperatures of Heated Domestic Equipment*. (London, U.K.: British Standards Institution).
- BSI (1983) *Medical Information on Human Reaction to Skin Contact with Hot Surfaces*. (London, U.K.: British Standards Institution).
- Bull, J.P. (1963) Burns. *Postgrad. Med. J.*, **39**, 717–723.
- EN 563 (1994) Temperatures of Touchable Surfaces, *Ergonomics Data to Establish Temperature Limit Values for Hot Surfaces*. (Brussels: CEN).
- Fanger, P.O. (1970) *Thermal Comfort*. (Copenhagen, Denmark: Danish Technical Press).
- Fiala, D., Havenith, G., Brode, P., Kampman, B. and Jendritsky, G. (2012) UTCI – Fiala Multi-node model of human heat transfer and temperature regulation. *Int. J. Biometeorol.*, **56**, 429–441.
- Haslam, R.A. and Parsons, K.C. (1987) A comparison of models for predicting human response to hot and cold environments. *Ergonomics*, **30**, 1599–1614.
- Houghten, F.C. and Yagloglou, C.P. (1923) Determining equal comfort lines. *J. Am. Soc. Heat. Vent. Eng.*, **29**, 165–176.
- ISO 7243 (1989) Hot environments – Estimation of the heat stress on working man, based on the WBGT-index (wet bulb globe temperature). (Geneva, Switzerland: International Standards Organisation).
- ISO 7726 (1998) Thermal environments – Instruments and methods for measuring physical quantities. (Geneva, Switzerland: International Standards Organisation).
- ISO 7730 (2005) Ergonomics of the thermal environment – Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. (Geneva, Switzerland: International Standards Organisation).
- ISO 7933 (2004) Ergonomics of the thermal environment – Analytical determination and interpretation of heat stress using calculation of the predicted heat strain. (Geneva, Switzerland: International Standards Organisation).
- ISO 8996 (2004) Ergonomics of the thermal environment – Determination of metabolic rate. (Geneva, Switzerland: International Standards Organisation).
- ISO 8886 (1992) Ergonomics – Evaluation of thermal strain by physiological measurements. (Geneva, Switzerland: International Standards Organisation).
- ISO 9920 (2009) Ergonomics of the thermal environment – Estimation of thermal insulation and water vapour resistance of a clothing ensemble. (Geneva, Switzerland: International Standards Organisation).
- ISO 10551 (1995) Ergonomics of the thermal environment – Assessment of the influence of the thermal environment using subjective judgement scales. (Geneva, Switzerland: International Standards Organisation).
- ISO 11079 (2007) Ergonomics of the thermal environment – Determination and interpretation of cold stress when using required clothing insulation (IREQ) and local cooling effects. (Geneva, Switzerland: International Standards Organisation).
- ISO 11399 (1995) Ergonomics of the thermal environment – Principles and application of relevant International Standards. (Geneva, Switzerland: International Standards Organisation).
- ISO 12894 (2001) Ergonomics of the thermal environment – Medical supervision of individuals exposed to extreme hot or cold environments. (Geneva, Switzerland: International Standards Organisation).
- ISO 13732-1 (2006) Ergonomics of the thermal environment – Methods for the assessment of human responses to contact with surfaces – Part 1: Hot surfaces. (Geneva, Switzerland: International Standards Organisation).
- ISO TS 13732-2 (2001) Ergonomics of the thermal environment – Methods for the assessment of human responses to contact with surfaces – Part 2: Human contact with surfaces at moderate temperature. (Geneva, Switzerland: International Standards Organisation).
- ISO 13732-3 (2005) Ergonomics of the thermal environment – Methods for the assessment of human responses to contact with surfaces – Part 1: Cold surfaces. (Geneva, Switzerland: International Standards Organisation).
- McIntyre, D.A. (1980) *Indoor Climate*. (London, U.K.: Applied Science Publishers).
- Parsons, K.C. (2014) *Human Thermal Environments*. (Boca Raton, FL, USA: CRC Press), ISBN 978-1-4665-9599-6.
- Smith, T.A. and Parsons, K.C. (1987) The design, development and evaluation of a climatic ergonomics knowledge based system. In: *Contemporary Ergonomics 1987*, E.D. Megaw (ed.), (London, U.K.: Taylor & Francis Group), pp. 257–262.



---

# 25 Assessment of the Visual Environment

*Peter A. Howarth*

## CONTENTS

|                                                   |     |
|---------------------------------------------------|-----|
| Introduction.....                                 | 678 |
| Scope of This Chapter .....                       | 678 |
| How Light Affects the Visual System .....         | 679 |
| Spectral Aspects .....                            | 679 |
| Temporal Aspects .....                            | 680 |
| Spatial Aspects .....                             | 680 |
| Parameters of Interest and Their Assessment ..... | 682 |
| Amount of Light .....                             | 683 |
| Illuminance.....                                  | 683 |
| Illuminance Units .....                           | 685 |
| Luminous Intensity (Point Sources).....           | 685 |
| Luminous Intensity Units .....                    | 685 |
| Luminous Flux Units.....                          | 686 |
| Luminance (Extended Sources) .....                | 686 |
| Luminance Units .....                             | 686 |
| Reflectance .....                                 | 687 |
| Reflectance Units.....                            | 687 |
| Flicker .....                                     | 688 |
| Colour .....                                      | 689 |
| Glare.....                                        | 691 |
| Discomfort Glare.....                             | 691 |
| Disability Glare .....                            | 693 |
| Directionality of Light, and Modelling .....      | 695 |
| Daylight.....                                     | 697 |
| Measuring the Visual Environment.....             | 697 |
| Standards and Guidelines .....                    | 697 |
| What Measures Are Needed?.....                    | 697 |
| An Example.....                                   | 698 |
| Average Illuminance Within the Room .....         | 698 |
| Desk Illumination .....                           | 699 |
| What Other Measurements Are Needed? .....         | 699 |
| Assessment of the Environment .....               | 699 |
| Checklist .....                                   | 700 |
| Light Levels.....                                 | 700 |
| Surfaces .....                                    | 701 |
| Glare.....                                        | 701 |
| Temporal Aspects .....                            | 701 |

|                                                           |     |
|-----------------------------------------------------------|-----|
| Chromatic Considerations.....                             | 701 |
| Spatial Considerations.....                               | 701 |
| Users.....                                                | 702 |
| Non-Visual Considerations of the Visual Environment ..... | 702 |
| Appendix: Spreadsheet for Modelling Calculation.....      | 702 |
| References and Information Sources .....                  | 703 |

## INTRODUCTION

### SCOPE OF THIS CHAPTER

Vision provides us with more information than all of our other senses combined. In the context of work, an increasing number of jobs involve the use of visually-displayed information and, as a consequence, the environmental conditions necessary to optimise the eyes' performance are of paramount importance. This involves more than simply providing a large quantity of light, as the more light there is the higher the cost and possibly the greater the light pollution.

Traditionally, the visual environment has been considered as consisting of two elements: the objects being viewed (such as paper copy, the road, instruments and dials) and their lighting (such as daylight, room lights or task-specific lights). However, the increasing use of self-luminous sources, such as televisions, computer screens, tablets and mobile phones has led to a reconsideration of this viewpoint for the indoor environment. This is because the visual characteristics of a self-luminous object are affected differently from those of an illuminated piece of paper, or material. Indeed, while an increase in illumination on paper will raise the luminance difference between text and background, it will reduce the contrast between the two when they are displayed on a screen.

There are many different aspects of the visual environment to consider when trying to produce the conditions necessary for good visual performance, not all of them immediately obvious. Consider an example: visual displays incorporating liquid crystal alphanumeric characters are to be installed in a self-service petroleum pump. What visual considerations are there? Some spring to mind immediately, such as character readability and the positioning of the display so that it can be seen by drivers of different heights, sitting or standing. But the relevant elements of the visual environment are not simply those related directly to the visual task in isolation from its surroundings. Other questions need to be asked: Is the display going to be used both during the day and night? If so, how is it going to be lit, and will supplementary lighting be needed? What is known of the spectral characteristics of the light that will fall on the display? Will it be lit by monochromatic sodium road lighting, and what account needs to be taken of this in the use of colour in the design? If supplementary lighting is going to be used, how should it be positioned to avoid producing veiling reflections in the display, and will discomfort be produced from glare? Is the display going to be covered with either glass or plastic, and if so what reflections will be produced in this cover by artificial lights, by car headlights, or by the sun as it crosses the sky? Clearly the issues involved in assessing the visual environment are not restricted to the ambient light levels!

This chapter discusses conditions of the visual environment that may be of concern to the ergonomist, and indicates why and how they may need to be assessed. We can think of the visual environment in three categories – *spectral* (to do with the wavelengths of light), *temporal* (to do with changes over time) and *spatial* (to do with changes over space). Of necessity, the chapter is general in nature: the aim is to provide the reader with sufficient information to enable him or her to approach a wide range of problems with an adequate understanding of the key issues. With the background information provided here, texts dealing with specific issues – such as the SLL (Society of Light and Lighting) and the IESNA (Illuminating Engineering Society of North America) Guides, or national/international standards such as BS EN 12464-1: 2011 – should be more readily understood. Finally, further information on the issues considered here, along with full references, can be found in standard texts such as Boyce's *Human Factors in Lighting* (2014).

## HOW LIGHT AFFECTS THE VISUAL SYSTEM

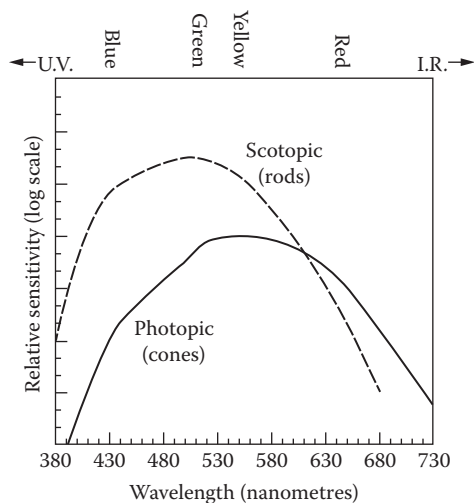
### SPECTRAL ASPECTS

Light is a small portion of the electromagnetic radiation spectrum. This spectrum includes radio waves, microwaves, ultraviolet radiation, infrared radiation and x-rays. What is special about light is that we can see it. While this seems an obvious statement, there is an important fact to be gleaned from it – namely, ‘light’ is defined by the human visual system, and not by the light source. If you cannot see it, it is not light. The sun, for example, radiates a wide range of the electromagnetic spectrum, some of which passes through the atmosphere and some of which is filtered out by it. However, the limits of what we call light are those of the human eye (approximately 380–740 nm [ $10^{-9}$  m]), not those of the sun. The spectral sensitivity of the eye, termed the  $V_\lambda$  function, is shown as a solid line in Figure 25.1. A photometer measuring *light* will, in effect, measure the total amount of radiation present over the spectrum as weighted by this function. This is done by passing the radiation through filters, so that the combination of the filter and the radiation sensor has the same response characteristics as the human eye (C.F. the weighting for sound and hearing).

The effectiveness of electromagnetic radiation in producing the sensation of vision depends upon how sensitive the eye is to the wavelengths present.

The *spectral characteristics* of the light (i.e. the relative amount of radiation at each wavelength) give it its characteristic appearance (i.e. its colour), and the weighted sum of the radiation present at each wavelength determines the *amount of light* present.

A potentially troublesome point to note here is that the spectral sensitivity of the eye is not totally independent of the light level. The eye has two kinds of photoreceptor in the retina, and these have different operating ranges. One kind are called ‘cones’, of which there are three types (providing us with the sensation of colour) the other kind are called ‘rods’. At normal light levels the cones are the active photoreceptors and the rods are essentially saturated (their photopigment is highly bleached). However, at lower light levels (for example, around dusk) there is not enough light for cones to operate, and rods become the active photoreceptors. The spectral sensitivity of the rod system differs from that of the normal cone system, and the change in sensitivity of the eye as the light level changes from



**FIGURE 25.1** The spectral sensitivity of the eye for daytime light levels (solid line) known as the  $V_\lambda$  function, and the sensitivity at night-time light levels, away from artificial sources known as the  $V'_\lambda$  function. The colour names describe the sensation experienced when radiation of that single wavelength is viewed (c.f. a rainbow).

photopic (cone) to scotopic (rod) levels is termed the 'Purkinje shift'. Purkinje, a Czech physician, noticed in 1825 that while red and blue paint on signposts looked the same brightness during daylight, at night the blue looked much brighter than the red. There are two aspects of the Purkinje shift to note from Figure 25.1: going from cone vision (solid curve) to rod vision (dashed curve), (1) the peak of the function moves from 555 to 507 nm and (2) the overall sensitivity of the eye increases.

From the functions shown in Figure 25.1, we can see that the absolute sensitivity of rods and cones is very similar at long wavelengths. This explains why long wavelength light is used to 'preserve' dark adaptation (e.g. on ship's bridges at night) – at these long wavelengths, when the light level is high enough for cones to operate, there is much less bleaching of rods than there would be at shorter wavelengths.

Other sources of variation in spectral sensitivity exist, and these include differences between stimulus conditions and between people. To standardise light measurement, a number of 'standard observers' with defined spectral sensitivities for particular stimulus field sizes, field positions and adaptation levels have been specified by the CIE (Commission Internationale de l'Eclairage). The 'standard observer' is a useful concept for defining units and standards and although individuals differ from it, these differences are generally too small to be of practical significance. Unless otherwise indicated, the calibration of a light meter will use the normal photopic function which is derived from the solid curve shown in Figure 25.1.

The three cone types present in the normal eye each have a different spectral sensitivity (the  $V_{\lambda}$  graph of Figure 25.1 is a composite function for the whole eye). A given wavelength will stimulate each cone type by a different amount, and it is the *relative* stimulation of each that gives rise to the sensation of a particular colour. The individual colours of a rainbow are seen because each wavelength produces a different ratio of cone stimulation. However, the same ratio (and hence the same colour appearance) can be produced by different combinations of wavelengths. A mixture of long wavelength (red) light and short-middle wavelength (green) light can provide the same relative stimulation as a middle wavelength (yellow) light and, in appearance, the two will be indistinguishable.\* The colour-rendering properties of these two light sources will be very different, however. Under the mixture a rose will look red and a leaf will look green, whereas under the single light both would look yellow.

## TEMPORAL ASPECTS

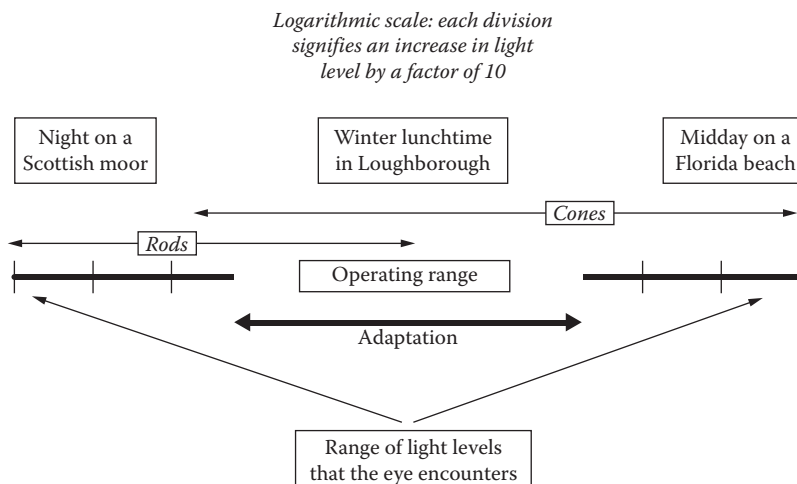
The *temporal characteristics* of the light source may be of interest when artificial lights are used, either because flicker is noticeable in the environment or because of interactions between the light source and the equipment being used. The visual system integrates light over a finite period of time, and a light flickering faster than the integration time will be perceived as steady rather than flickering. This integration time varies with light level, and so modern films, for example, appear to be continuous rather than flickering, even though their frequency is below the maximum that humans can detect under optimal conditions. Fast changes are not the only ones to consider, however, and within an interior one needs to consider changes over the day as the sun moves over the sky.

## SPATIAL ASPECTS

The *spatial characteristics* of tasks and lighting are the third aspect of the visual environment considered here. The human visual system is wonderfully adaptable, as shown in Figure 25.2, and operates over a wide range of ambient light levels. One reason for this is that the eye has two different

---

\* This is a *metameric* match: two objects (in this case, the light sources) appear to be identical in colour even though their spectral composition differs.



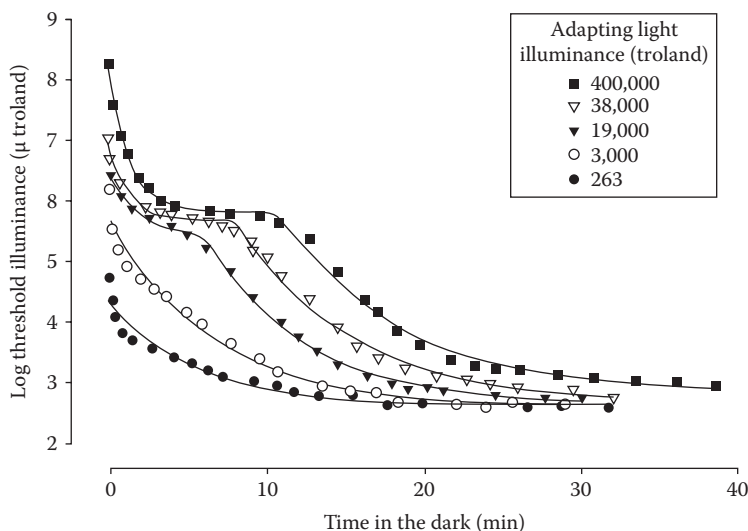
**FIGURE 25.2** The adaptability of the human visual system.

types of photoreceptors, rods and cones, each of which operates over a different portion of this overall range. At light levels above  $3 \text{ cd m}^{-2}$ , the *photopic* range, cones play the major role in vision. The visual system is unable to detect light at levels below  $10^{-6} \text{ cd m}^{-2}$  and in the *scotopic* range, between  $10^{-6}$  and  $10^{-3} \text{ cd m}^{-2}$ , only rods are functioning. The area between  $10^{-3}$  and  $3 \text{ cd m}^{-2}$  is called the *mesopic* range wherein both rods and cones are operating. The mesopic range corresponds roughly to dusk, at which time colours are hard to make out.

The other reason for the width of the range is that the sensitivity of each type of photoreceptor can vary with the ambient conditions. At any one point in time the eye will be sensitive over a restricted portion (its 'operating range') of its full range, and this operating range can change. Most people have experienced the situation in which they enter a darkened cinema, from the foyer, to discover that they can see very little. However, after having sat for five minutes or so, their vision has improved dramatically as they have adapted to the darker conditions. The time it takes to adapt depends upon both the starting and finishing ambient conditions, and so while it is commonly quoted that *full* dark adaptation may take half an hour, in most practical circumstances, the time taken to adapt to the (darker) ambient conditions will be appreciably less.

Figure 25.3 shows the time taken to reach a steady state of total dark adaptation following exposure to adaptation light of different intensities. Consider the uppermost curve (filled squares) which has two clear branches. The first branch represents the sensitivity of cones, which are initially the more sensitive of the two photoreceptor systems. After 10 min, the rod sensitivity has increased to such an extent that these photoreceptors are now more sensitive than the cones, which brings about the second branch. From this curve, we can deduce that if the eye is not placed in total darkness, it will actually reach its 'steady-state' adaptation level much sooner – well before 30 min – and, indeed, if the ambient light level is sufficient for cones to operate (e.g. in the cinema), then the complete change in the sensitivity will take <10 min.

Adaptation to lower ambient light levels is not a rapid process, whereas adaptation to higher ambient light levels generally only takes a second or two. So although dark adaptation and light adaptation both play the role of changing the eye's sensitivity to match the ambient conditions, they employ different physiological mechanisms. Light adaptation 'wins out' over dark adaptation in environments where there are extremes in spatial variation, and a large range of light levels within the environment are to be avoided, because visual performance will be degraded at the extremes. This effect can be seen in large rooms when sunlight shines through a window, providing lots of light for the areas near the window but leaving the areas away from



**FIGURE 25.3** The time taken to reach a steady state of total dark adaptation following exposure to adaptation light of different intensities. A troland is a unit of retinal illuminance. (Based on Hecht, S. et al., *J. Gen. Physiol.*, 20, 831, 1937.)

it relatively gloomy. Here, without supplementary lighting, the details in the shadows may be lost, because the adaptation state of the eye leaves it insensitive, with the details being below threshold.

On the other hand, some luminance (and chromatic) variation, which can occur across a room, across a workplace and at different heights from the ceiling, can be advantageous, and not just because of the pleasantness of the change. Directional lighting can facilitate tasks such as inspection. Also, lighting may be deliberately arranged within a room so that features stand out; this can be for safety (highlighting fire extinguishers and exits) as well as for aesthetic reasons.

## PARAMETERS OF INTEREST AND THEIR ASSESSMENT

There are three main reasons why measurement of the visual environment is usually undertaken:

1. *Health and safety.* The lighting of an area should be adequate to ensure that people can live safely, in comfort, and it should not in itself be a health hazard. Measurement of the visual environment can provide information as to whether or not these criteria are met.
2. *Visual performance.* There is a large body of knowledge telling us how visual performance is affected by variables in the visual environment, such as illuminance and contrast. Assessment of these variables can provide information about the expected performance in the location considered.
3. *Aesthetic reasons.* A pleasant environment is conducive to well-being, and will usually result in less stress and better task performance.

Laws, standards and guidelines exist to provide information about the requirements for the first two reasons. The third is often a matter of personal preference, although many guidelines and recommendations for aesthetically pleasing lighting seem to have good general support.

The following sections discuss the various parameters which might be relevant and provide background information about each; the *taking* of measurements is discussed further at the end of the chapter.



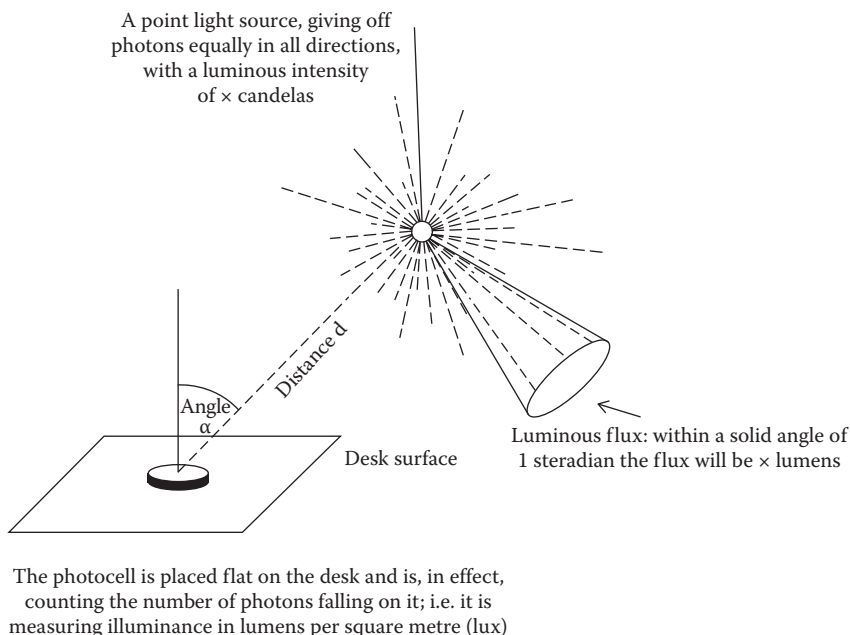
## AMOUNT OF LIGHT

In the recent past, measurement and specification of light has been complicated by the use of different units in different countries, and by different groups of people within the same country. The following text uses currently accepted SI units. Alternative units will still occasionally be encountered, for example in instructions for older instruments, and in older texts, and so descriptions of some previous units and their conversion factors have been included here.

The adaptability of the human visual system is such that we are very poor at making absolute judgements (but very much better at making *relative* judgements) and the eye itself is not a particularly trustworthy meter. Hence, light levels are usually measured objectively (although their subjective appearance should not be ignored), and Figure 25.4 illustrates the different concepts involved. There are two distinct aspects to consider. The first is the question of how much light there is present to ‘light things up’ – how much light is falling on this book, or how much light goes into your eye. This is termed *illuminance*. The second is how much light is being given off by something, and this is termed *luminance* (or luminous intensity if the source is small). The difference between illuminance and luminance can be illustrated by closing this book (after reading the next sentence!). The amount of light falling on the book is constant – the illuminance does not change – but overall, the page reflects more light than the book cover, and so the page has a higher average luminance.

### ILLUMINANCE

If we took a photocell which faithfully recorded every photon which landed on it and placed it on a desk, then the photocell output would depend upon the number of photons falling on it, no matter where they had originated, which is termed the *illuminance* on the desk surface. Illuminance measures tell you how much light is falling within the photocell catchment area, and nothing else. This amount of light could have been produced by a single bulb, a bank of fluorescent tubes or daylight – the source is irrelevant here; the measurement simply tells you how much light there is.



**FIGURE 25.4** The measurement of illuminance.

Illuminance varies with the position of the source. Imagine you had a very small light ‘bulb’ hanging from a wire above a desk. If the lamp was radiating equally in all directions (and no light was reflected from the walls or ceiling), the number of photons falling on the photocell would depend upon three factors:

- The area of the photocell’s collecting surface
- The distance from the source to the photocell
- The angle between the desk surface and the light source

Consider these in turn.

The larger the photocell area, the greater the number of photons falling on it; make the catchment area twice as big and twice the number of photons are captured. However, the number of photons captured *per unit area* would be constant. Unless there is variation across the desk surface, the size of the collecting area of the photocell is irrelevant. We can think of an illuminance measurement as being a measure of how many photons per unit area are falling on the desk, and the unit used for measurement (‘lumens per square metre’ which for convenience is termed ‘lux’) reflects this fact.

The further away the photocell is from the source, the fewer photons will fall on it, and we can easily determine the quantitative relationship between the source distance and the illuminance value. Imagine that the photocell had an area of 1 cm<sup>2</sup> and was directly below the tiny light bulb at a distance of 1 m. Because the source radiates equally in all directions, the amount of light falling on the photocell is proportional to the ‘solid angle’ made by the photocell at the source. The solid angle is the area of the photocell divided by the square of the distance from it to the source, and is measured in steradians (sr). So here the 1 cm<sup>2</sup> photocell at 100 cm distance would subtend  $1/100^2 = 1 \times 10^{-4}$  sr.

The point to note here is that the solid angle varies inversely with the *square* of the distance, and from the previous deduction, this means that the number of photons falling on the photocell varies inversely in proportion to the *square* of the distance from it to the source. Because the illumination on a surface is expressed in terms of the amount of light falling on it per unit area, the first law of illumination (the inverse square law) follows: the illuminance on the surface varies with the square of the distance between the surface and the source.

To illustrate the way in which *illuminance* depends on the distance from the source to the object being illuminated, imagine a flashlight with a diverging beam shone directly at a wall from a couple of inches away. If you walked away from the wall, two things would happen: the circle of light on the wall would increase, and the amount of light falling at *any point* on the wall would decrease. Double the distance from the torch to the wall and you increase by four the area of wall illuminated. This is because you both double the width of the illuminated patch and you double its height, and so you decrease the illuminance (i.e. the amount of light per unit area of the wall) by four as the same number of photons are spread over an area four times larger. Treble the distance and you reduce the illuminance by a factor of 9, quadruple it and there is a 16-fold reduction and so on.

The aforementioned examples have assumed that the photocell was positioned perpendicularly to the source. What if the photocell were angled – which is what would happen if it was placed on an inclined document holder rather than flat on the desk – with the light source directly above it? Here the amount of light per unit area falling on the photocell would be less than if it were perpendicular to the source, and the second law of illumination (the cosine law) is that the illuminance varies with the cosine of the angle that the photocell makes with the perpendicular. If we start with the photocell directly perpendicular to the source, the illuminance will be at a maximum value. If we now rotate the photocell so that its face is no longer exactly perpendicular to the source, the illuminance on the face will be reduced as fewer photons will be captured. When the photocell is rotated by 60°, for example, the illuminance will be reduced by a factor of two, because  $\cos 60^\circ = 0.5$ . Ultimately when the photocell is at 90°, the illuminance will be zero ( $\cos 90^\circ = 0$ ).

TABLE 25.1  
Examples of Recommended Illuminance Values

| Condition                                                                                                                | Recommended value<br>(lux) |
|--------------------------------------------------------------------------------------------------------------------------|----------------------------|
| Rarely visited locations, with limited perception of detail required, for example, railway platforms                     | 50                         |
| Continuously occupied areas, with limited perception of detail required, for example, railway waiting rooms, ticket hall | 200                        |
| General offices                                                                                                          | 500                        |
| Vehicle painting touch-up; typesetting                                                                                   | 1000                       |
| Assembly work (fine detail, e.g., watchmaking)                                                                           |                            |
| Inspection of extremely fine detail (e.g., small instruments)                                                            | 2000                       |
| Autopsy dissection table                                                                                                 | 5000                       |

Source: From BS EN 124644-1 *Lighting of Indoor Workplaces*, 2011.

Illuminance Units

The amount of light falling on a given surface area, the luminous flux per unit area, is the illuminance at a point on the surface. The SI unit is the ‘lux’ which is one lumen per square metre ( $\text{lm m}^{-2}$ ). That is to say, when luminous flux (see the following text) of 1 lm is spread over a surface area of 1  $\text{m}^2$ , the illuminance is 1 lx. An alternative name, not in common use, for the lux is the ‘metre candle’, while the ‘phot’ is the illuminance when 1 lm is spread over 1  $\text{cm}^2$ .

The *foot-candle* is a unit still sometimes encountered, especially on old illuminance meters; it is an illuminance of 1  $\text{lm ft}^2$ , and is equal to 10.76 lx.

The *troland* is a unit used in vision research; when an eye with a 1  $\text{mm}^2$  entrance pupil views a surface with a luminance of 1  $\text{cd m}^{-2}$  (see following text), then the retinal illuminance is 1 troland.

For illustration, Table 25.1 provides some representative indoor illuminance values recommended by the SLL.

LUMINOUS INTENSITY (POINT SOURCES)

In the previous section, we considered the light source to be very small, and to give off light equally in all directions. If this source was so small that it effectively had no area, it would be called a ‘point source’. In fact, this is a theoretical concept as all real sources have a finite area to them. However, the idea of radiation being emitted from a single point equally in all directions is a useful one in explaining the concepts of luminous intensity and luminous flux. The more intense a source is, the more light it gives off in all directions: luminous intensity is an attribute of the light source. Now consider a cone with a point source at its centre: the more intense the source, the more photons there would be within this cone, in other words, the greater ‘luminous flux’ there would be within the cone (see Figure 25.4). So luminous flux describes the light itself, and not the source.

An ideal point source gives off radiation equally in all directions. Although real sources can sometimes be assumed to have negligible area (consider the angular subtense of a motorcycle headlight viewed from a mile away), the property of directionality does not generally hold. Interior and exterior light fittings are often designed with reflectors or shades to produce a certain pattern of light, and the ‘hot spots’ of a headlight are a good example of this.

Given that the luminous flux varies with direction, it follows that the luminous intensity of a real (as opposed to a theoretical) source also varies with direction.

Luminous Intensity Units

Luminous intensity is a measure which describes the amount of light emitted by a point source. The SI unit is the *candela* (cd). This basic unit of light is actually defined in terms of the emission of a black body at the freezing point of platinum, 2040 K.

## Luminous Flux Units

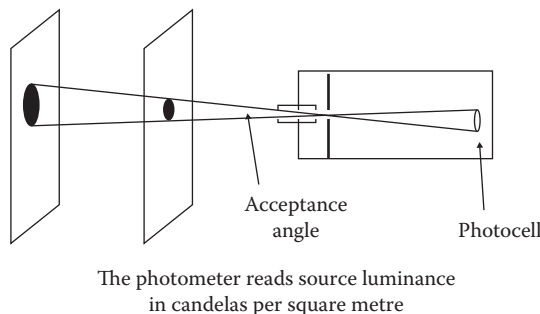
When light is emitted from a point source, the ‘density of photons’ within a cone centred at the source is termed the luminous flux. The number of photons within the cone will be a function of both the source intensity and the cone size (i.e. its ‘solid angle’) and the SI unit of luminous flux is the *lumen* (lm). This is defined as the luminous flux emitted through a solid angle of 1 sr from a point source of intensity 1 cd. A point source with a luminance of 1 cd will emit overall a total of 12.57 (i.e.  $4\pi$ ) lm.

## LUMINANCE (EXTENDED SOURCES)

In practice, one generally needs to consider extended sources rather than point sources. Usually, a fluorescent tube, a light fitting or some other source such as the sky subtends an appreciable angle at the eye, and cannot be considered to be infinitesimally small. Hence, luminance is specified in candelas per square metre of the source rather than in candelas. To take the measurement, a light meter with a small acceptance angle is used which will collect light within, for example  $1^\circ$ ,  $0.5^\circ$  or  $1'$ , depending on the instrument. It makes no difference how far away the source is, as long as it is larger than the meter’s acceptance angle. Consider a luminance meter pointed at a wall, collecting and recording all of the light within its acceptance angle. If we move the meter closer to the wall, more photons will be collected by the photocell from any given point. However, because the acceptance angle of the meter is fixed, the portion of the wall from which photons are being collected by the photocell is also reduced. The increased number of photons collected from each point is balanced by the decreased area of the wall within the meter’s acceptance angle, and so the luminance reading remains constant. Luminance, unlike illuminance, is independent of the measurement distance (Figure 25.5).

## Luminance Units

Luminance describes how much light is emitted from a surface, and the SI unit is the *candela per square metre* ( $\text{cd m}^{-2}$ ). (This unit in the past has been called a ‘nit’ but, to the chagrin of schoolchildren everywhere, this name is now deprecated. Despite this, it has recently been used in descriptions of the light output of screens and projectors.) The other unit that may be encountered is the *stilb*, which is  $1 \text{ cd cm}^{-2}$ .



**FIGURE 25.5** The measurement of luminance. The meter will, in effect, count photons that enter within its ‘acceptance angle’. The further the meter is from a source (e.g. a wall), the fewer photons will be captured from any point on the source (c.f. the inverse square law), but the larger will be the area covered. If the meter is moved away by a factor of 3, then the area covered by the acceptance angle will be three times wider and three times higher. This ninefold increase is balanced by the ninefold decrease in the number of photons counted from any point on the source, and so luminance is independent of testing distance.

## REFLECTANCE

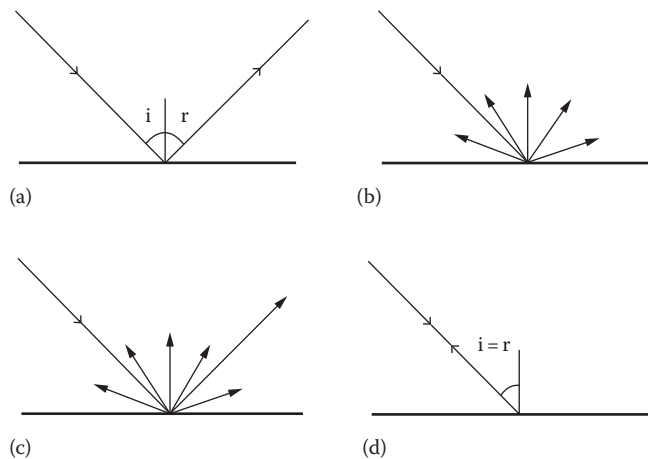
Illuminance and luminance are closely linked. The amount of light falling on a point on the wall is its illuminance, and the amount of reflected light coming back from the wall is its luminance. If all of the light that fell on the wall was reflected, then the values of the illuminance and the luminance would be the same, using appropriate units. If some of the light was absorbed, then the values would differ. The *reflectance* of the wall may be found by comparing the illuminance and the luminance values, as explained in the following text.

An important point to consider before going further is the *directionality* of reflections. A perfect matt white surface will reflect light equally in all directions, irrespective of the direction of the ambient light, and is called a perfectly *diffuse* reflecting surface (Figure 25.6b). On the other hand, a highly faithful reflector, like a mirror, will only reflect light at the same angle to the normal that the ambient light makes (Figure 25.6a). Most surfaces lie somewhere between these two extremes, with a higher luminous flux occurring at the angle of reflection than elsewhere (Figure 25.6c). (Try moving this book and see if you can detect specular reflection on the paper from any light sources.) Finally, if a surface reflects light back along its own path, it is termed a retro-reflector.

The greater the proportion of unwanted specular reflection from a surface, the more likely a person is to experience annoyance, discomfort and degraded visual performance. Where specular reflection is likely to be a problem, the positioning of light fittings must take this into account: areas of relatively high luminance should be avoided at places where they might be reflected into the eye. Where new equipment is used in a previously designed room, such as new screens in an old office, or tablets, the aim should be to allow for adjustability in the equipment so that specular reflections from the screen can be reduced.

### Reflectance Units

Irrespective of its angle of incidence, all of the light falling on a *perfect diffuse* reflecting surface is emitted equally in all directions. No light is lost, and it is intuitively appealing to have units which are defined in such a way that the values for luminance and illuminance are identical here. This is the rationale behind the old (now deprecated) luminance units ‘apostilb’ and ‘foot lambert’. When



**FIGURE 25.6** The four reflections: (a) specular – a perfect specular reflector (e.g. a mirror) reflects light at an angle ( $r$ ) equal to the angle of incidence ( $i$ ), (b) diffuse – a perfect diffuse reflector reflects light equally in all directions, irrespective of the angle of incidence of the illuminating light. (The arrow lengths represent the proportion of light reflected), (c) mixed – most surfaces will reflect some light in all directions, but will reflect a larger proportion at the angle of reflection, (d) retro-reflective – a retro-reflector reflects light back along the path it came. This is achieved by appropriate configuring of the surface.

the illuminance falling on a perfect diffuse reflector is 1 lx, the luminance of the surface will be 1 *apostilb* (asb). It follows that if the surface does not reflect all of the light, then the luminance in apostilbs will be equal to the reflectance of the surface, multiplied by the illuminance in lux. The same rationale applies for imperial units, and with an illuminance of one foot-candle (i.e. 1 lm ft<sup>-2</sup>), the luminance of a perfect diffuse reflector will be one *foot-lambert*. (It follows that 10.76 asb = 1 foot-lambert.)

The accepted SI luminance unit is the candela per square metre, and for a perfectly diffuse reflecting surface, there is a simple relationship between it and the apostilb:

$$1 \text{ cd m}^{-2} = \pi \text{ asb.}$$

So if the illuminance on a wall of reflectance 0.8 is 220 lx, then the luminance of the wall will be

$$(220 \times 0.8) = 176 \text{ asb} \quad (25.1)$$

or

$$\frac{(220 \times 0.8)}{3.14} = 56 \text{ cd m}^{-2}. \quad (25.2)$$

The first calculation is more straightforward, so why is the use of the apostilb deprecated? The reason is that in practice, most surfaces are *not* perfectly diffuse, but rather have some specular component. As a consequence, the wall luminance will change with viewing position and because the value of illuminance is fixed (and the same for all viewing angles), the reflectance will vary, depending upon the observer's position.

Instead of using the reflectance of the wall to determine the wall luminance, a quantity called the 'luminance factor' is used. The luminance factor is the ratio of the luminance of a reflecting surface, viewed in a given direction, to that of a perfect white diffusing surface identically illuminated. If the reflecting surface is itself a perfect diffuser, then the value of the luminance factor is the same as the reflectance, is independent of the viewing position and cannot be greater than one. On the other hand, if the surface does have specular reflections, then the luminance factor will vary with viewing position and at the angle of reflection, it could be greater than one. Hence,

$$\text{Luminance (cd m}^{-2}\text{)} = \frac{\text{illuminance (lx)} \times \text{luminance factor}}{\pi}.$$

The reflectance value of a surface is normally found for 'light', that is, the whole of the visible spectrum. It is also possible (although not normally needed in practice) to determine the reflectance values for different wavelengths. If this were done, then it would be seen that different coloured walls have different reflectance spectra. For example, a wall which looks red might reflect a large proportion of long-wavelength light, but little medium- or short-wavelength light, while one which looks green might reflect a large proportion of medium-wavelength light but little long- or short-wavelength light.

## FLICKER

Flicker, noticeable rapid fluctuations in light level, can be a serious problem in artificial environments. The visual system is very sensitive to flicker, which is related to its sensitivity to movement – both of which are changes in the visual scene over time. The *critical fusion frequency* (CFF) is the maximum temporal frequency (in Hz) at which flicker can be detected. Under normal interior photopic conditions, the highest CFF encountered is around 60 Hz, and hence, the typical 100 Hz flicker of fluorescent lights in the United Kingdom and Europe (120 Hz in the United States) is



undetectable. Like visual acuity, CFF declines with luminance and at low light levels, such as in the cinema, flicker of a given frequency is less detectable. As a consequence, the flicker of a film projected at 25 frames per second will not be noticed.

Unfortunately, objective measurement of flicker is not simple because it requires rapid-response equipment, normally available only to a lighting specialist. Subjective assessment of flicker is, however, much more feasible, and both the area of noticeable flicker and the degree of noticeable flicker can be adequately assessed by descriptive means. Also, because the periphery of the eye is more sensitive than the central area to flicker, subjective assessment may actually be a more relevant method. Hence, when dealing with flicker, the precise circumstances under which it is seen, such as the luminance and the position of the source in the visual field, should be noted.

In considering the subjective assessment of flicker, we should take a lesson from noise assessment (see Chapter 26 of this book). In the same way that the psychological effect of noise can be independent of the sound level (think about the dripping of a tap), flicker can have an annoyance or a distractive effect out of all proportion to its physical magnitude. A subjective assessment of the flicker should not only consider the physical aspects of the stimulus, such as the perceived flicker strength, but should also evaluate the psychological effect that the flicker is having on the person. The *positive* side to flicker is that, because it is very attention getting, its use is a very good visual method of conveying warning information.

Two further aspects need to be considered. First, some people are especially sensitive to flicker. Epileptics are an extreme example, and for them, flicker (particularly at frequencies around 10 Hz) can provoke seizures. This problem was highlighted some 10 years ago in Japan, when the “Pokemon incident” happened, in which a large number of schoolchildren were hospitalised following exposure to flashing lights on a cartoon TV show which provoked photosensitive epilepsy fits (Ishiguro et al., 2004). Generally, however, if a person has reached adulthood without experiencing these fits, then they will not do so throughout life. Second, flicker which is not visually detectable may still affect parts of the visual system. The human retina responds to flicker at high frequencies (over 100 Hz) even though the light appears steady, and no flicker is seen. It remains to be determined whether other parts of the human visual system are affected by high flicker frequencies, and whether performance or comfort is affected.

## COLOUR

Colour is a sensation which can be a powerful tool in the design of displays, and it may be defined in terms of hue, saturation and lightness (which, in the Munsell system are termed hue, chroma and value respectively [Boyce, 2014]). Hue describes the perceived colour, for example, red; saturation describes how pale or how dark the colour is (or how much ‘white’ there is – add white to red and you get pink); lightness is how bright it is.

At low light levels, when rods are the active photoreceptors, we do not experience the sensation of colour. However, our eyes contain three different type of cones photoreceptor, and although each will absorb light over a wide range of wavelengths, they each have different sensitivities over the spectrum. The sensation of colour (e.g. from a rainbow) then comes about because light of a given wavelength will stimulate each of these three cone types to a different extent, in the same way that colour televisions produce a range of colours by varying the luminance ratio of the blue, green and red pixels. Optimally, we are able to discriminate between colours as close as 2 nm in wavelength although our ability to discriminate between desaturated (pale) colours is poorer. Most visual stimuli are not composed of the single wavelengths found in the different parts of the rainbow, but instead are polychromatic (containing a range of wavelengths). The sensation of colour comes about in exactly the same way, however, as this depends upon the relative amount of stimulation of the three cone types, and not the actual wavelength content of the stimulus. Consequently, as already noted, the sensation of yellow can be achieved by showing the eye light of a single wavelength or by showing it an appropriate mixture from lamps that look red and lamps that look green.

Colour vision is defective in some individuals. This can be either congenital, in which case the problem for those people is generally confusion between reds and greens, or acquired (e.g. through diabetes) in which case the confusions are generally between blues and yellows. The congenital problem is genetically determined and is caused by a defective gene on the X chromosome. As it is recessive, a woman (who is XX) with one defective gene will still have normal colour vision, unlike a man (XY) with a defective gene whose colour vision will be abnormal.

As well as differences between people arising from their eyes, the perceived colour of an object can be altered by changing the lighting. The colour of an object is determined both by the spectral composition of the light source, and by the spectral reflectance properties of the object and its surround. In this way, two objects, like a shirt and tie, can appear identical in colour under one set of lights (a metameric match) but can be noticeably different from each other under another set of lights. The eye is easily fooled under these circumstances. However, while changes in colour appearance like this do occur, the adaptability of the visual system is such that most objects will retain their correct appearance under a wide variety of light sources. This 'colour constancy' can be seen outdoors when a cloud passes over the sun – although the spectral composition of the light falling on the grass and the trees changes drastically, the colours of the objects do not seem to change.

In the northern hemisphere, daylight coming from a northern sky is broad-band and is the reference illuminant used for colour-vision testing. Different lamps are compared with this standard as far as their colour-rendering properties are concerned, and the CIE have devised a 100-point scale, the 'colour-rendering index' (CRI) for lamps. As a generality, the higher the value of the CRI, the better the lamp performs (e.g. incandescent lamps may have a CRI of 99, an artificial daylight fluorescent lamp could have a CRI of 93, while a white fluorescent lamp has a CRI of 56). However, these are overall values for the lamps, and a lamp with a high score does not necessarily perform well over the entire spectrum, although it should give good rendering of most colours.

The fact that different light sources have different colour-rendering properties is reflected in the requirements published in standards such as EN 12464-1 for different environments. It is likely that the colour-rendering properties of the light sources will have been considered by the lighting designer in an environment where it is important, for example, where colour discrimination or colour matching is included in the tasks performed in that location. This occurs more often than casual thought would suggest; for example, text and instructions are often colour-coded. There is an important point to stress here: colour coding is valueless if the ambient light does not reveal the colour differences. Consider sodium light, which is monochromatic. In this light, the three cone types must always be stimulated in the same ratio. All objects viewed under sodium street lamps appear to be the same colour, yellow, and depending on how much of the light they reflect, some are darker than others. Hence, the colour coding of road signs has to be revealed at night by means of supplemental lighting.

Good colour rendering is not always necessary, and other criteria may take precedence. In some situations (such as industrial exteriors and warehouse interiors), the lights may have a simple safety function. Good colour discrimination is not required, for example, if the lighting only has to reveal the presence or absence of objects. Here the cost of running the lights may be a more important criterion than their pleasantness.

In an environment where people have to spend a large proportion of their working time, the colour-rendering properties of the lights, or combination of the lights, play an increasingly important role in determining the pleasantness and acceptability of the environment. Lamps with poor colour rendering are generally considered to provide a less pleasant environment than those with good rendering. Of particular importance here is the appearance of skin tones: the better the appearance of skin, the more subjectively preferable is the light. Although 'warm' lamps are generally considered preferable to those with a 'cool' appearance, lamps with a narrow spectral band (such as high pressure sodium) generally provide an unacceptable environment for long-term working. Also of interest here, and a factor where little research has been performed, is how people performing different occupations prefer different spectral combinations. Draughtspeople and designers, for example, may tend to use cool, blue-white light, while office work is more commonly performed under warmer, yellower light.

Colours will often have specific culturally based meaning associated with them (e.g. red = stop, green = go)\*. Here assessment of the use of colour involves measuring the performance associated with the colour rather than the colour itself. Also, visual performance can sometimes be enhanced by using colour to provide a contrast between a task and its surround. Unfortunately, while the concept of colour contrast has some intuitive meaning, (most people would consider the colour contrast between red and blue to be greater than that between yellow and green), as yet, there is no widely accepted objective metric by which it can be evaluated. At present, this assessment has to be made on a subjective basis.

## GLARE

Two distinct lighting-related visual problems, discomfort and a reduction in visual performance, have been grouped together under the heading of 'glare'. These problems have in common the fact that they are the negative consequences of light. Although the different forms of glare may occur simultaneously, they are essentially independent because they do not have the same underlying physiological mechanism.

## DISCOMFORT GLARE

When a portion of the visual field has a much higher luminance than its surround, a feeling of discomfort may occur around the eyes and brow. While this effect has been studied empirically for many years, the physiological mechanisms involved are still unknown. The facial and the iris muscles have been suggested as the site of the discomfort, but possibly neither is involved. When a glare source is turned on, people tend to wrinkle their brow and partially close their eyelids; concurrently, the iris sphincter muscle contracts and the pupil constricts. Whether these are simply *responses* to the increased light and the associated discomfort, or whether they *cause* the discomfort, is not known.

Although the mechanism of discomfort glare is unknown, the conditions under which discomfort occurs have been well established for a number of years. The early work of Hopkinson, Petherbridge and Guth (Hopkinson and Collins, 1970) revealed that

1. Discomfort *increases* with an increase in the luminance of the glare source and an increase in the angular size of the glare source at the eye.
2. Discomfort *decreases* with an increase in the luminance of the background and an increase in the angular position of the glare source relative to the line of sight.

By definition, discomfort is subjective, and discomfort glare is not easily quantified. A given physical configuration of lights will not only give rise to different reported amounts of discomfort from different people, but also to different reported amounts of discomfort from the *same* person on different occasions. Subjective assessment in this situation is not particularly reliable. On the other hand, the *physical* parameters of different lighting configurations are, in theory, easy to determine. If it is known how the aforementioned four factors (glare source luminance, size, position and background luminance) interact, it should be possible to measure aspects of the environment and determine on an objective scale how good or how bad the environment is. This is the rationale behind the various glare indices established in different countries – they say little about how an individual will respond, but they do allow an objective evaluation to be made of the lighting configuration. Calculation of the

---

\* Nature has developed colour coding, as seen in the colouring of wasps and bees: yellow and black hoops warn of a stinging insect (or an insect impersonating one). While 'yellow' is a human sensation (the insect itself is not yellow/black, but rather we sense it that way), we can infer from the colour coding that whatever might otherwise prey on these insects have some form of colour vision. Similarly, the use of colour in camouflage depends upon the colour vision of the observer: two objects may be metamERICALLY matched to a person with normal colour vision, yet to someone with a colour vision defect (or a normal person looking through a coloured filter), they will look different. Because of this, people with colour vision defects have been used in the past to 'spot' camouflaged objects.

glare index involves determining the luminance, size and position of the glare source(s). Full details for the calculations involved may be found in SLL and IESNA publications (see the information list at the end of this chapter) and in CIE publications 117–1995 and 190–2010.

Despite the apparent validity of the laboratory studies carried out in carefully controlled experimental conditions, the evaluation of the glare indices in practice is complicated. While the luminance and size of the glare source(s) are easily measured, there are real difficulties in determining the value of both the background luminance and the glare source position. First, the luminance of different walls and different parts of the ceiling will vary, and the problem arises of what value constitutes the ‘true’ background luminance. In practice, an ‘average’ value has to be estimated to represent the ambient conditions. Second, there is not one unique glare index for one room position, but rather one glare index for each position in the room and each position of the eyes.

A parameter mentioned earlier is the angle between the line of sight and the glare source, and it follows that as someone looks around a room, the glare index will vary. So while a head and eye position may be specified for measurement purposes, for example, the person may be assumed to be looking down at their desk, or perhaps to be looking straight ahead, each position is arbitrarily chosen and is not one which the person necessarily adopts or has problems with.

Not surprisingly, there is a poor correlation between subjective reports of discomfort and the glare index. The advantage of the index is that it does provide an objective description of the environment. The disadvantage of the index is that this figure does not in itself describe well the subjective discomfort of an individual subjected to that glare.

An alternative way of evaluating glare is to use an ‘expert observer’ approach, where an individual experienced in glare assessment can evaluate the degree of discomfort they feel in a certain situation. The very first step that should always be taken is to initially determine whether there is a problem by shielding the suspected glare source(s) and seeing if there is an improvement in comfort. If so, then you can safely assume that there was some discomfort present in the first place. This can then be evaluated more precisely. A number of descriptive scales have been produced, perhaps the most useful of which is that proposed by Hopkinson (Hopkinson and Collins, 1970). Four criterion points were used, each of which describes a transitory position of subjective feeling:

Just perceptible  
Just acceptable  
Just uncomfortable  
Just intolerable

The scale should include an end point of imperceptible (zero), and if necessary it can be extended by allowing descriptions between the criterion points (e.g. not quite acceptable, but not yet uncomfortable). This will produce a nine-point scale. An alternative method is simply to rate the amount of discomfort on a rating scale, with defined anchor points having specific, known, rating values, for example, a rating scale where imperceptible = 0, just uncomfortable = 5 and intolerable = 10. (See Chapter 4 for general discussion of rating scales.)

There are a number of ways to reduce discomfort glare. Consider in turn the four factors mentioned previously. Reducing the glare source luminance will reduce the discomfort. This solution may not always be feasible, because the source could be providing illumination for a different part of the room. Increasing the angle between the viewer’s line of sight and the glare source, by moving either the source or the person, will decrease the discomfort. Another, less obvious, solution is to *increase* the background luminance. It seems counterintuitive that increasing the amount of light present reduces the discomfort, but this manoeuvre reduces the contrast between the source and the background. A practical way of doing this is by painting the walls and ceilings so that they reflect more light, thereby increasing the ambient light level. Finally, the solid angle of the glare source can be reduced by shielding the source. Changing the luminaire fitting to one with a narrower cut-off angle will decrease the visible extent of the source.

These solutions are not necessarily independent. For example, by reducing the size of the glare source, the illuminance it provides for some other purpose may be decreased, necessitating an increase in the source luminance. In the same way, a bare light bulb or a luminaire with a small fitting or shade may be improved by increasing the size of the fitting. Although the glare source solid angle is increased, the source luminance is lowered, with a consequent reduction in discomfort. These are all solutions which are imposed on the person by outside manipulation of the light sources. Where feasible, it is preferable to allow individuals control over their immediate visual environment, with directional local lighting, and in this way enable them to position the light sources in such a way that they are comfortable.

As a final point, the long-term effects of small amounts of discomfort glare are unknown. It is reasonable to suppose that someone working all day under conditions where glare is just perceptible will have a build-up of discomfort over the day. The need for improvement here is greater than in the situation where someone is briefly exposed to 'just uncomfortable' glare once or twice per day. Similarly, someone's idea of how tolerable glare is will depend upon what they are doing, and how interested and involved they are in it. A person watching television may choose to have a low ambient light level but a bright screen – environmental conditions which in other circumstances would be considered intolerable. The point to remember here is that the glare assessment cannot be considered in isolation from the context in which it is made.

## DISABILITY GLARE

Visual performance depends crucially upon contrast. We cannot usually see stars during the day-time, in a cloudless sky, not because they have moved to the other side of the world, but because the contrast between them and the sky is too low. While the physiological mechanisms involved in discomfort are not understood, the ways in which an extraneous light source can affect visual performance are quite clear, and occur in one of two ways. A glare source can either act on the task, as in the example of the stars, reducing the contrast between an object and its background, or it can act on the eye. These two mechanisms are quite distinct, even though they both reduce visual performance, and so are classified as 'task disability glare' and 'ocular disability glare', respectively.

Task disability glare can occur because of a discrete reflection, such as the specular reflection of a light source from the surface of a computer screen. Here the luminance of both the object (the characters being viewed) and the background (the surrounding screen) are raised by the addition of the extra light, but the contrast is reduced. It can also occur because of a diffuse reflecting veil over the whole of the task, as is seen, for example, when a car windscreen mists up or when the sun shines on a bank ATM machine screen. The whole of the scene looks grey and washed out, and both luminance contrast and colour contrast are diminished. These examples have in common that the contrast between the object and the background is decreased, with a consequent reduction in object visibility. Hence to reduce the disability, one should raise the contrast between the task and the background.

Ocular disability glare affects the eye and not the visual task. It is seen, for example, when a car approaches at night with its headlights on full beam. The disability in this situation has two sources:

1. There is scatter within the eye, reducing the retinal image contrast.
2. The adaptation level of the eye is raised as the car approaches, and the eye is dazzled.

After the car has passed, it takes a little while for the eye to re-adapt to the ambient light level.

The need for the ergonomics/human factors (E/HF) practitioner to distinguish between 'task' and 'ocular' disability glare can be seen when one considers age-related changes to the eye. The media within the eye generally become more opaque with age, and, in particular, the lens tends to go 'cloudy'. Although these changes will have only a minor effect on task disability glare, the real problems are seen in the presence of ocular disability glare. This is because, for the older person, opacities within the media will scatter the light to a much greater extent than is found for a younger person.

Hence, an older person who is able to perform perfectly well under normal lighting conditions may be severely disabled in the presence of ocular disability glare- and the reason for this disability may not be apparent visually to a younger person. Both sources of disability can be present together, such as when an object is seen near a bright window and both the object and the eye are affected.

Ocular disability glare is usually found in circumstances where discomfort will also be present because in both cases, the eye itself is affected by the glare. Like discomfort, the disability can sometimes be reduced by raising the light level. On a city street, a car headlight, even when it is not dipped, causes less disability than the same light on an unlit road because the extra ambient light raises the eyes' adaptation level. In the city, the headlight is less bright in comparison with its surroundings, and we can take this reasoning further by considering that the same headlight would produce virtually no disability during daylight.

There is a further problem, encountered less often, where performance is reduced and this is where there is an area of low light as compared with the ambient light level to which the eye is adapted. The problem faced here by the visual system is exactly the same as that faced by a photographer encountering deep shadows. If the exposure is set manually for the ambient light, the details in the shadows are lost, whereas if the camera is set to record these details, everything else in the photograph will be overexposed. The eye's adaptation system acts like an automatic camera, but it is not possible to vary it in the way you can with a manual override camera. The automatic setting the eye will adopt is appropriate for the ambient light, and not for the shadows, and so details in the shadows are lost to the visual system.

An overall diagram classifying glare on the basis of the user is provided as Figure 25.7.

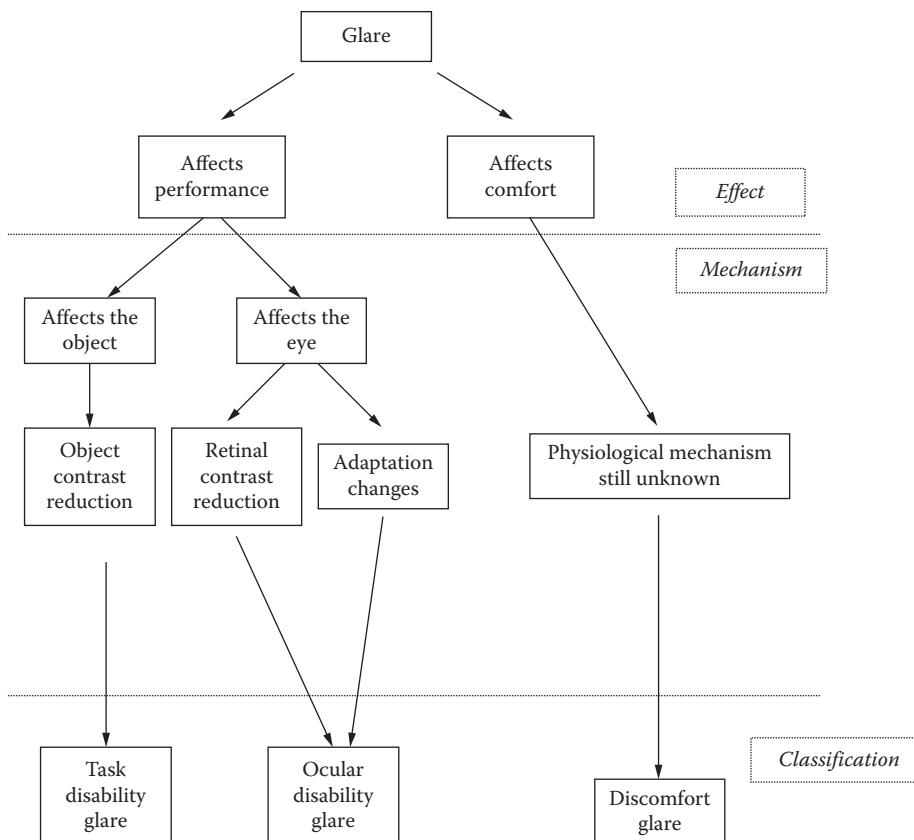


FIGURE 25.7 The classification of glare from a user's viewpoint.



## DIRECTIONALITY OF LIGHT, AND MODELLING

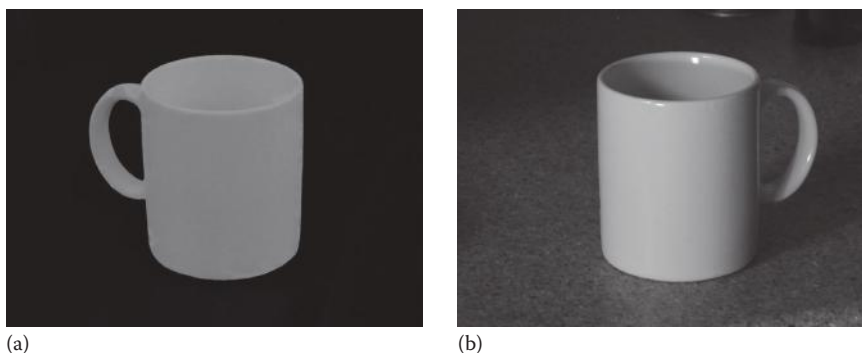
Lighting is said to look ‘flat’ if there is little modelling present (Figure 25.8). The amount of modelling can be evaluated by measuring the directionality of the light, and a single value can be obtained by combining two measures. One is the scalar illuminance; the other is the illumination vector (Figure 25.9).

First, define a reference direction within the room (such as the direction from the measurement point to the centre of the facing wall). This is needed so that the directionality of the light can be specified in comparison with this direction. Facing this direction, imagine that you have a cube in front of you. Hold the illuminance meter against each of the six imaginary faces, noting the illuminance falling on each. Averaging the values of the six cube faces gives the scalar illuminance. The vector illuminance is then calculated by constructing a vector diagram of the differences within each of the three pairs of opposing faces, as illustrated in Figure 25.9. Alternatively, the appendix to this chapter provides the formulas needed to construct spreadsheet, and once you have done this the vector and scalar illuminances are obtained when these results are entered into the spreadsheet.

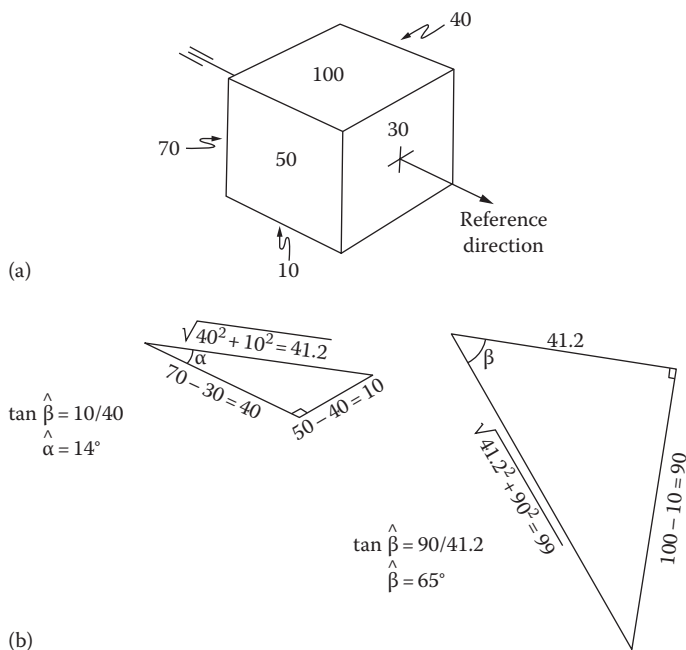
Combining the illumination vector value with the scalar value gives the vector/scalar ratio. A value of 0.5 is considered essentially shadow-free, while a value of 3.0 provides very strong modelling, with the details within the shadows often being invisible. Pleasant modelling of human facial features, which is generally an extremely important criterion in relation to the overall acceptability of the visual environment, occurs between values of 1.2 and 1.8. Other values are given in Table 25.2.

The directionality of lighting can also be used to reveal relief, and as a means of directing attention. For example, the inspection of textured items, such as cloth, may be enhanced by providing directional lighting which will cause shadows to be cast. The uniformity of these shadows can be easily assessed – defects showing up as an imprecision in their pattern. The elegant role that the lighting plays here is to facilitate performance by changing the nature of the visual task.

A further consideration in lighting uniformity is the illuminance distribution over a workplace. Variation in light level across a workplace will both differentially direct the attention to different areas, and will provide a more pleasing visual environment. For example, the whole surface of a desk may have an apparently satisfactory illuminance level according to published guidelines, but if there is no spatial variation across the desk, it appears uninteresting. This situation is greatly improved by providing slightly higher illumination on the area just in front of the person, where they put their immediate work. When the task is illuminated to a higher level than the surround, the appearance is much more interesting and pleasant.



**FIGURE 25.8** Object viewed under (a) flat and (b) directional light. (From Howarth, P.A., Role of vision in falls, in Haslam, R.A. and Stubbs D.A. (eds), *Understanding and Preventing Falls; Chapter 5*, Taylor & Francis Group, London, U.K., 2003. With permission.)



**FIGURE 25.9** Calculation of the scalar and vector illuminances: (a) The illuminance values are measured on each face of a cube suspended in space, and the scalar illuminance is their arithmetic average (50 lx). (b) A vector diagram can be constructed by first determining the difference between each pair of faces (90, 40 and 10 lx). Then the angle of the illumination vector in the horizontal meridian relative to the reference direction shown in (a) is determined as shown in the left-hand portion of (b);  $\tan \alpha = 10/40$  so  $\alpha = 14^\circ$ . The magnitude of this vector is the square root of  $40^2 + 10^2$  (the two measurements taken in the horizontal plane) which equals 41.2. The angle in the vertical meridian, relative to the direction just determined, can now be found, as shown in the right-hand portion of (b).  $\tan \beta = 90/41.2$  and so  $\beta = 65^\circ$ . The magnitude of the illuminance vector is the square root of the sum of the squares of these two sides: the square root of  $(41.2^2 + 90^2)$  which equals 99 lx. Answer: The illumination vector has a magnitude of 99 lx, in a direction  $14^\circ$  to the left of the reference direction, and  $65^\circ$  below this direction. The vector:scalar ratio = 99/50.

**TABLE 25.2**  
**Vector/Scalar Ratio and the Appearance**

| Vector/Scalar Ratio | Assessment of Appearance | Application                                          |
|---------------------|--------------------------|------------------------------------------------------|
| 4.0 (maximum)       |                          |                                                      |
| 3.5                 | Dramatic                 |                                                      |
| 3.0                 | Very strong              | Strong contrasts; details in shadows not discernible |
| 2.5                 | Strong                   | Suitable for displays; too harsh for human features  |
| 2.0                 | Moderately strong        | Pleasant appearance for distant faces (formal)       |
| 1.5                 | Moderately weak          | Pleasant appearance for near faces (informal)        |
| 1.0                 | Weak                     | Soft lighting for subdued effect                     |
| 0.5                 | Very weak                | Flat, shadow-free lighting                           |
| 0 (minimum)         |                          |                                                      |

Source: After Cuttle, C., *Lighting by Design*, Architectural Press, Oxford, U.K., 2008.

## DAYLIGHT

The need for windows in buildings, providing natural light, has come into question recently because of their cost in terms of heat-loss and energy conservation. The scientific evidence of a physiological need for windows is, at best, unproven; however, the psychological evidence is clear. A small, windowless room can easily be considered cell-like and restricting, while the presence of a window provides visual access to the outside world. Larger rooms are considered less restrictive, and in a well-controlled environment, the absence of windows becomes less important. However, in these windowless environments, the information, such as the time of day, and the variety provided by the changing outside light are still absent.

As well as the illuminance variation over the day, the provision of daylight can also provide spatial and spectral variation within a room, again decreasing the monotony of the environment. The illumination variation may be quantified as a change in the 'daylight factor' across a room. The daylight factor is the ratio of the illuminance from the skylight measured on a horizontal surface within the room to the illuminance from the skylight (not direct sunlight) measured on a horizontal plane which has an unobstructed access to the hemisphere of the sky. At different positions within the room, the daylight factor will vary and, if required, this variation may be assessed over a room. For interiors where the daylight factor variation is not large, such as rooms with skylights, or rooms which are not too deep, when the average daylight factor is 5% or greater, an interior will appear generally to be well day-lit. Also, if the illuminance from the sky is not known, the relative daylight factor at different positions within the room will give information about the spread of daylight within the environment.

The pleasantness of variation in spectral content and luminance level is not restricted to natural lighting. On the contrary, in some parts of the world, the design of mood lighting provides a considerable source of revenue for interior designers and fixture designers alike.

## MEASURING THE VISUAL ENVIRONMENT

### STANDARDS AND GUIDELINES

Lighting guidelines and standards come in two forms, either statutory, where certain legal requirements have to be met, or recommendatory, where the recommendations provided are given as examples of 'good current practice'. In neither case should the values given be assumed to be 'ideal'. For example in the case of a legal minimum requirement for a certain illuminance at a workplace, the *minimum* required is not necessarily the *optimum* value. Illuminance values which were considered 'good practice' 20 years ago are now, with the advent of more efficient lamps, often considered inadequate.

Most countries have their own standards and guidelines, and further specific values will not be presented here. Rather, the reference and further information sections contain details of documents which contain the necessary information.

### WHAT MEASURES ARE NEEDED?

An assessment of the visual environment is going to be needed in one of two situations. The first is a design situation where the concepts discussed previously will have to be considered in a somewhat abstract manner. The example given in the introduction, designing a petroleum pump display, is a case in point. The second situation is a more immediately practical case, where an environment already exists and has to be evaluated in some way. Generally, the assessment will be needed because problems are known to be present, and the appropriateness of the measures to be taken depends upon what these problems are.

The preceding sections have discussed the variety of factors of concern in the visual environment. Outside of the laboratory, many of these lend themselves to subjective assessment and flicker is a good example. Objective physical quantification of the frequency and modulation of flicker is complicated;

however, it is usually unnecessary. The appropriate question may be “Is there unpleasant flicker present?” – a question that has to be answered on subjective rather than on objective grounds.

In an assessment of an existing environment, the people who live and work in that environment are an important resource. As well as providing information about what environmental problems exist, their severity and the appropriateness of the measures being taken to alleviate them, people familiar with the environment can also provide information which reveals problems which are not immediately obvious. Problems which may occur at different times of the day, or only under certain circumstances (such as when an infrequently used piece of equipment is operating), fall into this category. Once the problem has been established, then the relevant measures can be taken.

**AN EXAMPLE**

Suppose you are approached by a client who feels that the average illuminance within an office is too low at the end of the day, and that some of the desks are inadequately lit. You are asked to make an assessment of the environment, and to suggest improvements. Off you go to the office armed with a luminance meter, an illuminance meter and a tape measure. What do you then do?

Consider this problem in the following four stages:

**Average Illuminance Within the Room**

The initial issues to be considered in determining the average illuminance within the room are ‘how many measurement points are needed, and where should the measurements be taken?’ The answers to these questions depend upon the size and dimensions of the room, and the accuracy required. The following method will give accuracy to within 10%.

The first step is to determine the dimensions of the room, including the ceiling height, and to sketch a scale plan of a cross-section of the room. The positions and dimensions of the luminaires should be marked on this sketch. Next, the ‘room index’ (an overall size – measure of the room parameters) is found from the ceiling height and the length of one wall. For a square room, it is numerically half the ratio of the wall length to the ceiling height. If the room is almost square, then take the measurement for the largest wall in determining the wall–ceiling ratio. The minimum number of measurement points can then be found from Table 25.3.

If the room cannot be considered as square, an extra calculation is involved. First measure the smaller wall length, and the ceiling height. From these figures, you can determine the minimum number of points needed within the square defined by the dimension of the *smaller* wall. This then tells you how far apart the measurements points need to be. By applying this value to the room *as a whole*, the required number of measurement points may be calculated. Take the example of a room which is 20 m × 30 m with a 4 m ceiling. If the room were 20 m × 20 m, then the wall–ceiling ratio would be 5, and from the aforementioned table, a minimum of 16 measurement points would be needed.

**TABLE 25.3**  
**Minimum Number of Measurement Points Needed for Different Sized Rooms**

| Ratio Wall Length: Ceiling Height | Room Index | Minimum Number of Points |
|-----------------------------------|------------|--------------------------|
| <2                                | <1         | 4                        |
| 2–4                               | 1–2        | 9                        |
| 4–6                               | 2–3        | 16                       |
| >6                                | >3         | 25                       |

So in the 20 m × 30 m room, which is 1.5 times as large, 24 measurement points would be needed to maintain the distance between measurement points.

Greater accuracy than 10% can be achieved by increasing the number of measurement points, and when time allows the more points taken the better. Doubling the number of points halves the error, and accuracy to within 5% can then be claimed.

Once you have determined the number of measurement points, the next step is to divide the plan of the room into a grid consisting of a number of squares each equal in area. If necessary, the number of measurement points should be increased to ensure that the grid is symmetrical. The idea is to take illuminance measurements at the centre of each of these squares. However, before doing this, check from the plan that the measurements are not going to be biased by having a disproportionate number of measurements taken directly underneath luminaires. The illuminance measurements should be taken in the horizontal meridian, at around desk height, without shadows being cast on the meter and the values should be recorded directly onto the sketch plan of the office. In this way, you will also have a visual indication of the illuminance variation over the room to use in conjunction with the arithmetic average for the room as a whole.

### Desk Illumination

Again, it is a good idea to sketch a plan of what is being measured, in this case, each desk. While doing this, note down what measurements need to be determined, and under what conditions they need to be taken. One position on the desk may be identified as being the location where the primary tasks are performed, with other areas being either where objects used less are placed, or simply storage space. If objects on the desk are inclined, in the way that a document holder may be, then the horizontal plane may not be the appropriate place to take the measurement and this should be noted on the sketch. Measurements should be taken with any occupants in place, so that any shadows usually cast by the person are present.

As well as the illuminance falling on the desk, the luminance of the various points of interest can be determined, as measured from the occupant's usual position. Knowing the illuminance on an object, and its luminance, the 'luminance factor' may be calculated if thought relevant.

### What Other Measurements Are Needed?

Depending on the circumstances, further measures could be appropriate. In the present example, the luminances and reflectances of the walls, floor and ceiling *should* be determined. Wall reflectance is usually important to the lighting of small rooms, but less so for large rooms where the area immediately adjacent to the wall is the only part of the room affected by light reflected from it. On the other hand, the larger the room, the greater is the importance of the ceiling reflectance. While a small room can have a satisfactory appearance with a darker ceiling, a large room needs to have a large proportion of the light incident on the ceiling reflected back into the room.

Optionally, a subjective 'expert observer' assessment of discomfort glare could be taken from each workplace, the relative daylight factor could be measured at each grid point, the contrast rendering factor could be established, measurements could be repeated to determine the variation over the day and so on. While these measures might all be outside the original brief, where time and enthusiasm permits, they enable a much more complete description of the workplace to be made and provide a sounder basis on which to make recommendations about appropriate changes.

### Assessment of the Environment

Now that the measurements have been taken, the values obtained can be compared with values known to represent good practice, as detailed in the SLL code. Here the recommendation for standard service illuminance (the mean illuminance over the lifetime of the lamps) for a general office

is 500 lx, with a minimum of 200 lx for any working surface. The code also gives recommendations which are applicable here for illuminance ratios:

1. The ratio of the minimum illuminance to the average illuminance over the task area should not be  $< 0.8$ . While a greater degree of non-uniformity is acceptable between task areas and areas of the room where no tasks are performed, the illuminance on areas adjacent to the tasks should still not be  $< 0.3$  of the task illuminance.
2. In an interior with general lighting, the ratio of the average illuminance on the ceiling to the average illuminance on the horizontal working plane should be within the range of 0.3–0.9.
3. In an interior with general lighting, the ratio of the average illuminance of any wall to the average illuminance on the horizontal working plane should be within the range of 0.5–0.8.
4. In an interior with localised or local lighting, the ratio of the illuminance on the task area to the illuminance around the task should be not more than 3:1.

By comparing the measured values with the aforementioned recommendations, the extent of any problem present in the office may be evaluated, and remedial action may be suggested. Here the extra measurements suggested may be helpful in differentiating between the available solutions. For example, if the wall and ceiling reflectances are acceptable, then the solution may lie in modification of the luminaires, whereas if these reflectances are low, the situation could be improved by painting the walls and ceiling.

This example is, of necessity, short and addresses a specific and relatively standard issue. Assessment of other visual environments will involve different circumstances and problems. The criteria by which the design of visual aspects of a car interior is judged will be rather different from those applicable in an office, for example. With this in mind, the final section of this chapter contains a checklist to indicate the range of problems which might be present in any visual environment, and the questions that may need to be asked.

## CHECKLIST

What are the problems?  
 What measures are needed?  
 What guidelines are available?

Consider each of the following if appropriate:

### LIGHT LEVELS

Luminance and illuminance:

Where is the light needed?  
 What variation is there?  
 Across the room?  
 Across the workplace?

Is supplementary light needed anywhere?

At any particular time of the day?  
 At any particular time of the year?  
 For particular purposes related to the immediate task?  
 For purposes unrelated to the tasks, for example, safety lighting.



**SURFACES**

What are the reflectances/luminance factors of the various surfaces?

- The walls
- The ceiling
- The floor

What are the reflectance ratios between them?

**GLARE**

Discomfort glare:

- Is it a problem?
- Subjective or objective assessment needed?
- Can it be relieved by modifying the environment?

Disability glare:

- Is it affecting performance?
- Can it be relieved by moving or shielding the lights?

**TEMPORAL ASPECTS**

Is flicker apparent:

- In the visual task itself, for example, a computer screen?
- In the environment, for example, from fluorescent tubes?

**CHROMATIC CONSIDERATIONS**

Is the colour rendering of concern?

- Is the colour rendering of the lights acceptable?
- Is the colour appearance of the lights acceptable?

Is colour discrimination a factor for concern?

- How good is the colour rendition of the present lights?
- Is supplementary light with better colour rendition needed for the task, to enhance colour discrimination? Is colour coding present, and if so are the lights adequate for the colours to be discriminated?

**SPATIAL CONSIDERATIONS**

Directionality:

- Is there a specific need for directional lighting?
- What variation is there over the room?
- What variation is there over the day?

How is the daylight supplemented by artificial light, and how does this vary over the day?  
Are shadows a problem under normal working conditions?

#### Highlighting:

Are there any features that need special consideration to increase their conspicuity?  
Are there any special-purpose needs, such as safety lighting with a back-up power supply?

#### USERS

Are there particular user groups with specific requirements?

#### NON-VISUAL CONSIDERATIONS OF THE VISUAL ENVIRONMENT

How can the environment be maintained under the desired conditions: What maintenance is needed?

What non-visual effects occur, for example, heat from luminaires?

As well as the following sources, information can be obtained from the bodies which produce standards in individual countries (e.g., British Standards, American ANSI standards, German DIN standards). Also, International Standards have been established in many areas by the International Standards Organisation (ISO).

#### APPENDIX: SPREADSHEET FOR MODELLING CALCULATION

---

Stand facing your 'reference direction'

What is the reading on the:

|                                                     |          |       |         |                                         |
|-----------------------------------------------------|----------|-------|---------|-----------------------------------------|
| Upper (top) face                                    | 100.00   |       |         |                                         |
| Lower (bottom) face                                 | 10.00    |       |         |                                         |
| Left face                                           | 50.00    |       |         |                                         |
| Right face                                          | 40.00    |       |         |                                         |
| Front face                                          | 30.00    |       |         |                                         |
| Back (towards you) face                             | 70.00    |       |         |                                         |
| Flow up/downwards =                                 | 90.00    |       |         |                                         |
| Flow to the left/right of the reference direction = | 10.00    |       |         |                                         |
| Flow to/away from the reference direction =         | 40.00    |       |         |                                         |
| Scalar illuminance =                                | 50.00    |       |         |                                         |
| Overall vector illuminance =                        | 98.99    | 14.04 | Degrees | To the right of the reference direction |
|                                                     |          | 60.95 | Degrees | Below this line                         |
| Vector/Scalar ratio =                               | 1.979899 |       |         |                                         |

To create this Excel spreadsheet, put

#### Cell

#### Cell Content

|    |                                         |
|----|-----------------------------------------|
| A1 | Stand facing your 'reference direction' |
| A2 | What is the reading on the:             |
| C4 | Upper (top) face                        |
| C5 | Lower (bottom) face                     |
| C6 | Left face                               |
| C7 | Right face                              |
| C8 | Front face                              |
| C9 | Back (towards you) face                 |

(Continued)

|       |                                                                         |
|-------|-------------------------------------------------------------------------|
| C12   | Flow up/downwards =                                                     |
| C14   | Flow to the left/right of the reference direction =                     |
| C16   | Flow to/away from the reference direction =                             |
| A18   | Scalar Illuminance =                                                    |
| A21   | Overall vector Illuminance =                                            |
| A24   | Vector/scalar ratio =                                                   |
| G21   | Degrees to the right of the reference direction                         |
| G22   | Degrees below this line                                                 |
| D12   | = (D4–D5)                                                               |
| D14   | = (D6–D7)                                                               |
| D16   | = (D8–D9)                                                               |
| D18   | = (D4 + D5 + D6 + D7 + D8 + D9)/6                                       |
| D21   | = $\text{SQRT}((D12 \times D12) + (D14 \times D14) + (D16 \times D16))$ |
| D24   | = (D21/D18)                                                             |
| D4–D9 | Values                                                                  |

*Notes:* To check it, input the values provided earlier, and the answers should correspond to the figures given.

## REFERENCES AND INFORMATION SOURCES

- Boyce, P.R. (2012). Illumination, in G. Salvendy (ed.), *Handbook of Human Factors and Ergonomics*, 4th edn., John Wiley & Sons, Inc., Hoboken, NJ. doi: 10.1002/9781118131350.ch24.
- Boyce, P.R. (2014). *Human Factors in Lighting* (CRC Press). This is the primary book for anyone interested in the area. The third edition was published recently. ISBN 9781439874883.
- BS 8206-2: 2008. Lighting for buildings. *Code of Practice for Daylighting*. This standard provides recommendations regarding design for daylight in buildings. It includes recommendations on the design of electric lighting when used in conjunction with daylight.
- BS EN 12464-1: 2011. Light and lighting: Lighting of work places. *Indoor Workplaces*. This has replaced the old BS8995:2002, British Standards Institution, London, U.K.
- BS EN 12464-2: 2007. Light and lighting: Lighting of work places. *Outdoor Workplaces*. British Standards Institution, London, U.K.
- BS EN 12665: 2011. Light and lighting. *Basic Terms and Criteria for Specifying Lighting Requirements*. British Standards Institution, London, U.K.
- CIE (Commission Internationale de l'Eclairage) publications: These also cover a wide range of issues, such as colour rendering, light measurement, and visual performance. Information on their availability can be obtained from lighting organisations of most countries (e.g., CIBSE, IES).
- CIE 117–1995. Discomfort Glare in Interior Lighting.
- CIE 190–2010. Calculation and Presentation of Unified Glare Rating Tables for Indoor Lighting Luminaires.
- Cuttle, C., *Lighting by Design*, Architectural Press, Oxford, U.K., 2008.
- Egan, M.D. (2001). *Concepts in Architectural Lighting*, 2nd edn., McGraw Hill, New York. A comprehensive book covering vision as well as lighting in an easily understood volume. A little dated now, but as yet there is no third edition.
- Hecht, S., Haig, C. and Chase, A.M. (1937). The influence of light adaptation on the subsequent dark adaptation of the eye. *The Journal of General Physiology* 20, 831–850.
- HSE (1998). Lighting at work. UK Health and Safety Executive publication Series Code HSG38 ISBN 9780717612321. Download available at: <http://www.hse.gov.uk/pubns/priced/hsg38.pdf>.
- Hopkinson, R.G. and Collins, J.B. (1970). *The Ergonomics of Lighting*, Macdonald, London, U.K. Outdated in parts, but still covers the basic principles well.
- Howarth, P.A. (2003). Role of vision in falls, in R.A. Haslam and D.A. Stubbs (eds.), *Understanding and Preventing Falls*, Chapter 5. London, U.K.: Taylor & Francis Group.
- IESNA *Lighting Handbook* (2011), 10th edn., Illuminating Engineering Society of North America, New York. This handbook is described by IESNA as “an indispensable reference for industry professionals ... known as the “Bible of Lighting.” It currently costs \$595 to non-members of IESNA.
- Ishiguro, Y., Takada, H., Watanabe, K., Okumura, A., Aso, K. and Ishikawa, T. (2004). A follow-up survey on seizures induced by animated cartoon TV program “Pocket Monster”. *Epilepsia* 45(4), 377–383.

- Pritchard, D.C. (1999). *Lighting*, 6th edn., Longman, London, U.K. This inexpensive paperback is a little dated now, and does not cover recent developments in lighting technology.
- SLL (CIBSE publications). In particular, note the SLL *Lighting Handbook and the Lighting Guides*. The SLL Lighting Handbook is designed to be complementary to the SLL Code for Lighting but to go beyond it in terms of applications and background information without getting into the fine detail of the Lighting Guides. These Guides cover a wide variety of issues, e.g., glare, lighting for computing, lighting guides for various applications such as engineering, libraries, hospitals. Information can be obtained from: <http://www.sll.org.uk/>.
- Smith, N.A. (2000). *Lighting for Health and Safety*, Butterworth Heinemann, Oxford, U.K. An excellent book, written at a level appropriate for students and practitioners. Primarily deals with the issues from a lighting viewpoint rather than from a human factors viewpoint.
- Society for Light and Lighting (2012). SLL code for lighting (used to be called the CIBSE Code for Lighting), see <http://www.sll.org.uk/>.
- Wyszecki, G. and Stiles, W.S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd edn., John Wiley & Sons, New York. A standard reference work on colour: Comprehensive, but not easy reading.

---

# 26 Auditory Environment and Noise Assessment

*Christine M. Haslegrave*

## CONTENTS

|                                                                           |     |
|---------------------------------------------------------------------------|-----|
| Introduction.....                                                         | 705 |
| Measurement of Sound .....                                                | 706 |
| Definition of Units.....                                                  | 708 |
| Comparison of Sound Levels .....                                          | 708 |
| Effects of Several Sources of Sound .....                                 | 710 |
| Measures of Noise.....                                                    | 710 |
| Subjective Measures.....                                                  | 711 |
| Weighted Measures .....                                                   | 711 |
| Empirical Measures.....                                                   | 711 |
| Frequency Analysis .....                                                  | 712 |
| Noise Measurement Procedures.....                                         | 713 |
| Instruments .....                                                         | 714 |
| Field Measurements .....                                                  | 714 |
| Laboratory Measurements.....                                              | 714 |
| Location of Microphone.....                                               | 715 |
| Identifying Noise Sources.....                                            | 716 |
| Measurement of Hearing Ability and Risk of Hearing Damage.....            | 716 |
| Measuring Hearing Ability.....                                            | 716 |
| Measuring Masking Thresholds.....                                         | 716 |
| Assessing the Risk of Hearing Damage .....                                | 717 |
| Measurement of Effects of Noise on Comfort and Performance.....           | 718 |
| Annoyance.....                                                            | 718 |
| Speech Intelligibility .....                                              | 719 |
| Design of Warning and Information Signals for Safety and Audibility ..... | 720 |
| Assessment of Personal Hearing Protectors .....                           | 721 |
| Evaluating Attenuation of Noise Levels.....                               | 721 |
| Evaluating Audibility of Communications.....                              | 722 |
| Concluding Remarks.....                                                   | 722 |
| References.....                                                           | 723 |

## INTRODUCTION

It is rare to experience silence in the modern world and our auditory environment is very complex. Sound can bring pleasure and information but unnecessary sound or too much sound is annoying, distracting and possibly harmful. Sound in our environment can be generated by transport of all forms, by equipment and machinery, and by people themselves. In the community, this is mainly

due to traffic, to neighbours and to televisions, radios and other domestic equipment. In an office, sound comes from people talking, telephones, computers and printers. In manufacturing industry, it comes mostly from motors, fans, pumps and compressors, moving machinery, contact between tool and workpiece, and resonant plates or housings. Traffic and machinery can also set up vibrations in the structure of the building. All these have to be considered when investigating the sources of unwanted or harmful noise and in finding ways of reducing the noise levels (through suppression or shielding) or of providing hearing protection.

It is not always possible to separate the positive and negative responses to sound. It is quite possible for a sound to be wanted by one hearer and unwanted by many others, and conflicts may arise between differing interests in both living and working environments. Ergonomists often need to evaluate auditory environments which may include sounds which are unwanted although they are not so loud as to be harmful, or sounds which are essential or enjoyable to one person while distracting or annoying to another. For this reason, we tend to distinguish between sound and noise. Kryter (1985) has defined noise in terms of its effects on people as 'audible acoustic energy that adversely affects the physiological or psychological well-being'. As he says, this is consistent with the usual definition of noise as 'unwanted sound'.

Noise levels therefore may be measured to assess whether there is a problem in terms of health, safety, communication, working performance, comfort or annoyance. Equipment may be needed to measure sound levels of individual sources and of the environment, exposure over a period of time, hearing ability and parameters relevant to risk of hearing damage. Correspondingly, an equally wide range of measures have been defined for such evaluations. This chapter describes the methods used for measuring sound (or noise) and some of the more common measures and criteria which have been proposed for assessing noise. The aim is to introduce various techniques used by ergonomists, but without attempting to cover specialist applications or the concerns of acoustic engineers who deal with noise control, architectural acoustics or design of communication equipment. Similarly, there is no intention of covering either the physics of sound or physiology of hearing, other than to introduce the necessary concepts and terms.

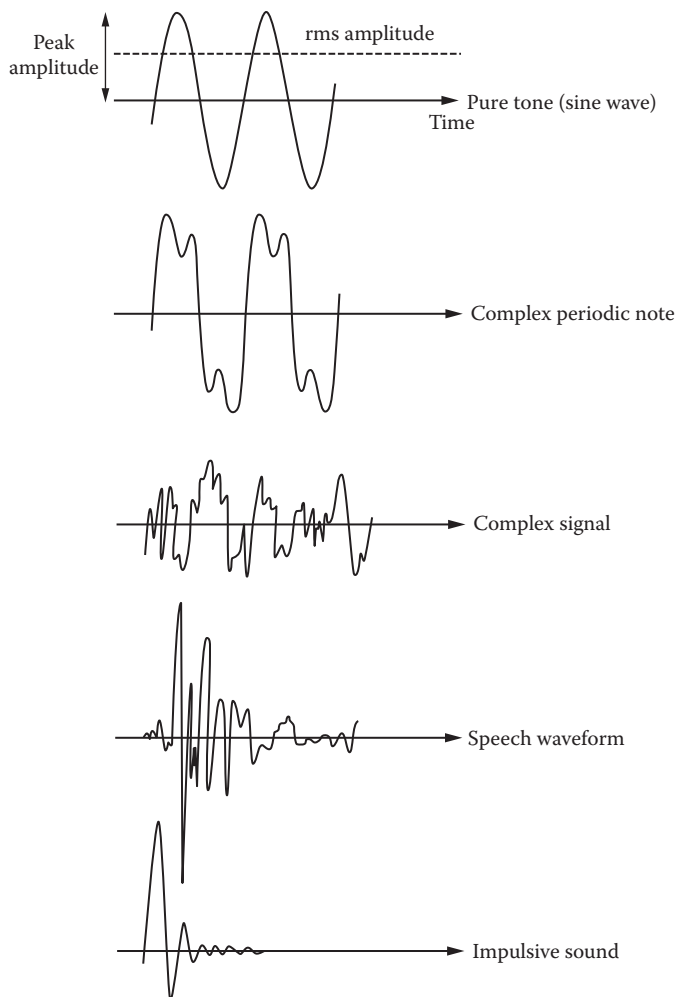
For a more detailed study of the ergonomic aspects of the auditory environment, recommended texts dealing with hearing, perception of sound and the effects of noise on work and health include Jones and Chapman (1984), Kryter (1985, 1994), Loeb (1986) and Sanders and McCormick (1992). Guidance on acoustic treatments and techniques for reducing noise levels can be found in technical noise control handbooks as well as in a series of case studies on the UK Health and Safety Executive website (<http://www.hse.gov.uk/noise/casestudies/>), some of which were originally published as a booklet (Health and Safety Executive, 1995) but which is now out of print.

## MEASUREMENT OF SOUND

Sound is a variation of pressure in the air (or in any other elastic medium) which the human sense of hearing detects. It is transmitted in the form of pressure waves as a series of compressions and rarefactions travelling outwards from the source of the sound. The speed of the wave depends on the medium, but in air at 20°C sound waves travel at a velocity of 344 m s<sup>-1</sup>. The amplitude of the sound wave is the fluctuation above or below the ambient air pressure. Our sensation of loudness however does not come directly from the air pressure but from the intensity of the sound wave, which is the rate at which sound energy is transmitted in the wave. This is defined in terms of the energy passing through a unit area in unit time (or sound power per unit area). Air pressure can be measured directly, and it is usual to measure sound levels in terms of sound pressure, and then to calculate sound intensity from these measurements, as described later.

Auditory stimuli at the ear result from the combined signals received from all sources in the environment. The waveforms (or variation in amplitude with time) of typical sound signals are shown in Figure 26.1. A pure tone (which might be obtained from a tuning fork or a vibrating string)





**FIGURE 26.1** Waveforms of various sound signals.

oscillates at a single frequency and can be represented as a sine wave, but most sounds are made up of complex signals containing many frequencies. Complex signals may be analysed into their component sine waves (each with a characteristic frequency, amplitude and phase) to understand the content of the signal. This process is known as frequency or Fourier analysis, giving a frequency spectrum that indicates the principal frequencies contained within the signal and their relative intensities (energies). When a sound is made up of frequencies covering most of the audible sound spectrum, it is known as white or broad-band noise. An impulsive signal (such as a door slamming or a hammer blow) is a single pressure pulse with a very fast rise time (around 25 ms or less) to the initial peak amplitude, followed by small pressure oscillations decaying over about 1 s.

The audible range of sound levels is enormous, from a quiet whisper to a warning siren (which is close to the pain threshold), representing a range of power in the sound signal of well over 1 to one billion, as can be seen in Table 26.1. Similarly, the ear is sensitive to a large range of frequencies – the audible range is approximately 20 Hz–20 kHz, with greatest sensitivity between 2 and 5 kHz. Most speech is between 300 and 700 Hz, with all vowel sounds below 1 kHz, but sibilant consonants may be higher than 5 kHz. Low frequency sounds below 20 Hz, usually termed infrasound, only seem to cause adverse effects (resonant vibration in

**TABLE 26.1**  
**Range of Intensity and Pressure of Audible Sound**

| Intensity ( $\text{W m}^{-2}$ ) | Pressure ( $\text{N m}^{-2}$ ) | Decibel Level | Example                     |
|---------------------------------|--------------------------------|---------------|-----------------------------|
| $10^{-12}$                      | 0.00002                        | 0             | Hearing threshold           |
| $3 \times 10^{-6}$              | 0.04                           | 65            | Conversation                |
| $10^{-4}$                       | 0.2                            | 80            | Town traffic                |
| $10^{-2}$                       | 2                              | 100           | Workshop                    |
| 3                               | 36                             | 125           | Jet at take-off (60 m away) |
| 100                             | 200                            | 140           | Pain threshold              |

chest, throat and nasal cavities and pain in the ear) at very high intensity (Kryter, 1994), and this frequency range is not usually considered when assessing the auditory environment.

### DEFINITION OF UNITS

Several units are used for sound measurement, and it is helpful to give some brief definitions to show the relationships between the various basic measures. A fuller description can be found in texts such as Kohler (1984).

Sound is described by intensity and by frequency. (The corresponding sensations in human hearing are termed loudness and pitch.) Intensity or power of the oscillations in the air is measured in watts per metre<sup>2</sup> ( $\text{W m}^{-2}$ ), while pressure is measured in newtons per metre<sup>2</sup> ( $\text{N m}^{-2}$ ); frequency is measured in hertz (Hz).

The total sound power of a sound source (such as a machine) can be measured and is expressed in watts (W).

In a sound wave emitted from a source, the power and pressure are related according to the following equation:

$$\frac{\text{Power}}{\text{Unit area}} = \frac{\text{Energy flow}}{\text{Unit area}} = \frac{\text{Pressure}^2}{\text{Density of air} \times \text{Speed of sound}}.$$

The sound power level is therefore proportional to the square of the sound pressure.

From the sound signals in Figure 26.1, it may be seen that the sound amplitude fluctuates rapidly over time, so the root mean square (rms) value of the pressure is the measure normally used. This is derived from the time average of the squared values of the instantaneous pressures over the duration of the measurement.

These are the units used for physical measurements of sound, but the auditory characteristics of the human ear are such that the corresponding sensations of loudness and pitch are not linearly related to the intensity and frequency of the sound. Measures of these sensations are therefore given separate units (which are defined later): loudness is normally measured in phons, and pitch is measured in mels.

### COMPARISON OF SOUND LEVELS

Sound levels are usually measured relative to other sound levels, or to a base reference level. For this, a unit called the decibel (dB) has been defined as the ratio of their intensities or powers. (The original unit was the bel, which represented a 10-fold increase in intensity, but this was found to be too large in practical applications.) The decibel is a logarithmic unit and thus corresponds quite closely to the non-linear response of the human ear.

**TABLE 26.2**  
**Decibel Scale**

| Ratio of Intensities (Powers)<br>of Two Sounds | Ratio of Pressures of<br>Two Sounds | Difference in Decibels between<br>the Two Sound Levels |
|------------------------------------------------|-------------------------------------|--------------------------------------------------------|
| 1                                              | 1                                   | 0                                                      |
| 10                                             | 3.16                                | 10                                                     |
| 100                                            | 10                                  | 20                                                     |
| 200                                            | 14.14                               | 23                                                     |
| 400                                            | 20                                  | 26                                                     |
| 600                                            | 24.49                               | 28                                                     |
| 800                                            | 28.28                               | 29                                                     |
| 1,000                                          | 31.6                                | 30                                                     |
| 10,000                                         | 100                                 | 40                                                     |
| 100,000                                        | 316                                 | 50                                                     |
| 1,000,000                                      | 1000                                | 60                                                     |
| 1                                              | 1                                   | 0                                                      |
| 1/10                                           | 1/3                                 | -10                                                    |
| 1/100                                          | 1/10                                | -20                                                    |

The base reference sound intensity is chosen as the power of a standard vibration in the air which is just on the threshold of hearing, or one thousand billionth of a watt per square metre ( $10^{-12}$  W m<sup>-2</sup>). Thus, a sound level of N dB is related to its intensity of I W m<sup>-2</sup> by

$$N \text{ dB} = 10 \log_{10} I/I_0,$$

where  $I_0$  is the reference level intensity at the threshold of hearing.

Given the relationship between sound intensity and pressure,

$$N \text{ dB} = 10 \log_{10} I / I_0 = 10 \log_{10} p^2 / p_0^2 = 20 \log_{10} p / p_0,$$

where  $p$  is the sound pressure level (SPL) and  $p_0$  is the reference level amplitude at the threshold of hearing (which is  $2 \times 10^{-5}$  N m<sup>-2</sup>).

The values of intensity, pressure and decibels for some typical sounds within the audible range are shown in Table 26.1. The relationships between the three scales are shown in Table 26.2. A sound which is 10 times louder than another has an intensity level which is greater by 10 dB. Since the decibel scale is logarithmic, a sound 100 times louder is said to be 20 dB louder. The smallest change in loudness that the human ear can discriminate (or difference threshold) is 1 dB, but in practice, the difference needed to recognise a sound above the background noise level is about 3 dB, which represents a doubling in intensity or loudness. A doubling of the SPL is equivalent to an increase of 6 dB in sound level.

Since the various units are related in this way (with loudness or sound power proportional to the square of sound pressure), a useful guide to differences in sound levels is given by

$$+3 \text{ dB} = 1.4 \times \text{SPL} = 2 \times \text{sound power level}$$

$$+6 \text{ dB} = 2 \times \text{SPL} = 4 \times \text{sound power level}$$

$$+20 \text{ dB} = 10 \times \text{SPL} = 100 \times \text{sound power level}$$

**TABLE 26.3**  
**Addition of Two Sound Levels in Decibels**

| Difference (dB) between Sounds | Add to the Louder Sound (dB) |
|--------------------------------|------------------------------|
| 0                              | 3                            |
| 1                              | 2.5                          |
| 2                              | 2                            |
| 4                              | 1.5                          |
| 6                              | 1                            |
| 8                              | 0.5                          |
| 10                             | 0                            |

### EFFECTS OF SEVERAL SOURCES OF SOUND

What then happens when two sounds are heard at the same time? How are their effects added? Sound levels are not directly additive when they are measured in logarithmic dB units. If two sounds are received at the ear simultaneously, the resultant is the sum of the energies ( $I_1 + I_2$ ) in the two sounds and so the combined sound power level becomes

$$\text{Combined sound power level (W m}^{-2}\text{)} = I_1 + I_2.$$

Thus, the decibel level is

$$\text{Combined sound level (dB)} = 10 \log (I_1 + I_2)/I_0.$$

The combined loudness of sounds from two or more sources can therefore be calculated by using this formula, but a simpler rule of thumb (accurate to 0.5 dB) is given in Table 26.3. If the two sounds are of equal intensity (difference 0 dB), their combined intensity is 3 dB higher than the intensity of either. If the difference between the sounds is 5 dB, the combined sound is approximately 1 dB higher than the louder of the two. If the difference between the sounds is 10 dB, the combined sound is approximately equal to the louder sound, and there is no noticeable difference in the intensity of the resultant sound. Examples are as follows:

Two sounds of 75 dB combine to give a sound level of 78 dB

Two sounds of 75 and 80 dB combine to give a sound level of 81 dB

Two sounds of 75 and 85 dB combine to give a sound level of 85 dB

As a practical consequence of this, the sound of a machine or piece of equipment can be measured even in the presence of background noise, provided that the background noise is at least 10 dB below the machine noise level. If the background noise is greater than this, a correction factor for the measured machine noise level can still be calculated as shown earlier. It is also useful to note that the ratio of two sound levels is calculated by subtracting the level in decibels: this gives a signal-to-noise ratio as the difference in decibel levels between the desired signal and the unwanted noise.

### MEASURES OF NOISE

It is obvious from this brief review that measurements of sound or noise levels will be influenced by the temporal, intensity and spatial characteristics of the signals. The temporal characteristics can include both the frequency spectrum and fluctuations in the overall sound level with time. A variety of measures have therefore been developed for specific purposes, and the most important are discussed in the following text.

## SUBJECTIVE MEASURES

Since the response characteristics of the ear are non-linear, both frequency and intensity affect our perception of loudness of a sound. For example, we do not perceive a sound level at a frequency of 50 Hz arriving at the ear to be as loud as if the signal frequency had been at 1000 Hz. Curves of equal subjective loudness can be plotted, giving the sound levels which provide constant perceived loudness at various frequencies and indicating the combinations of intensity and frequency which appear equally loud to a human listener. This shows that the 50 Hz tone must have a loudness of about 85 dB to give the same subjective loudness as a 1000 Hz tone at 50 dB.

Subjective loudness is measured in a unit called a phon. This unit is equivalent to the decibel at 1000 Hz and is also logarithmic. While the phon measures the subjective equality of sounds, a unit called the sone was defined to measure the relative loudness of sounds. One sone is defined as the loudness of a 1000 Hz tone of 40 dB (40 phons). A sound of 2 sones is one judged to be twice as loud. The phon and sone scales are related by the following formula:

$$\text{phon} = 40 + 10 \log_2 \text{sone}.$$

Every increase of 10 phons then doubles the loudness in sones, so that, for example, 50 phons is equivalent to 2 sones.

There is also a subjective measure of pitch, which corresponds to frequency for a pure sinusoidal tone and to the fundamental frequency for a complex waveform. Perceived pitch is given a unit called a mel – defined by a pure tone of frequency 1000 Hz at a SPL of 60 dB, which is said to have a pitch of 1000 mels. Any two tones separated by a given number of mels appear equally far apart in pitch, regardless of their frequency.

## WEIGHTED MEASURES

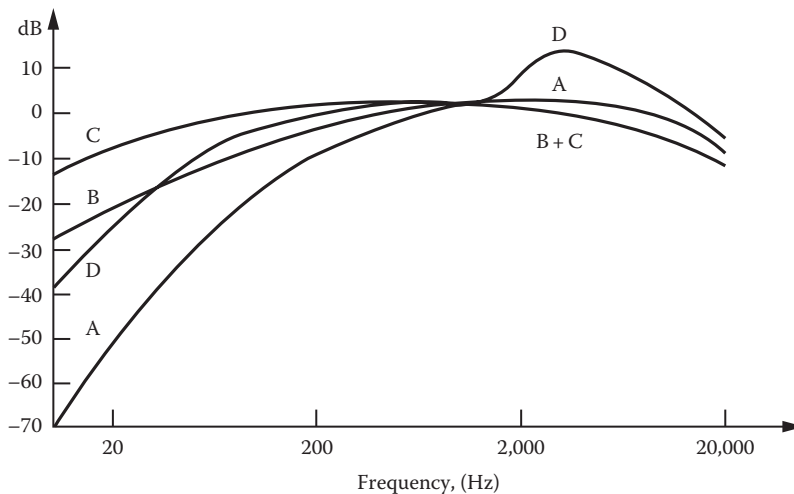
In view of the non-linearity of hearing response, most instruments are designed to measure sound levels on a decibel scale which is weighted to match the characteristic of the ear. Several scales have been developed for different purposes, and their response characteristics have been standardised internationally. The A-, B- and C-weighted scales were developed to match the responses for sounds of low, moderate and high intensity. The most commonly used is the dBA scale, which gives the best correlation with subjective tests of perceived loudness, and also with ratings of noise annoyance. The A-scale was in fact designed to match the 40 phon equal loudness contour.

This means that the actual measurement of the SPL is converted to a weighted dBA sound level, in accordance with the response characteristic shown in Figure 26.2 which takes account of the middle range of frequencies to which the human ear is most sensitive. The other scales shown in Figure 26.2 are less commonly used. The B-scale was designed to match the equal loudness contour at 70 dB, and the C-scale for a flat rating across the frequency scale. There are also D-weighted scales which were designed as equal noisiness scales for assessing aircraft noise. Unweighted sound levels are normally only used for frequency analysis.

## EMPIRICAL MEASURES

In most everyday situations, noise contains sounds from various sources, of differing frequencies and intensities, and also extending over different periods of time. In order to describe the noise environment, various statistical distribution measures are used. These are based on noise levels measured on a dBA scale. The most important are defined as follows:

*Equivalent level of sustained noise ( $L_{eq}$ ).* This is the average level of sound energy over a given period of time, which integrates all the fluctuating noises to represent them as an average steady level. It therefore takes account of short but high peaks.



**FIGURE 26.2** Frequency weighting characteristics for A-, B-, C- and D-weighted decibel scales.

*Median noise level ( $L_{50}$ ).* This is the noise level which is exceeded for 50% of the time.

*Background noise level ( $L_{90}$ ).* This is the 10th percentile level – the level that is exceeded for 90% of the time.

*Peak noise level ( $L_1$  or  $L_{10}$ ).* This is the 99th or 90th percentile level respectively – the level that is exceeded for 1% or 10% of the time.

All these indices can be used to give a measure of the total environment in, say, an office or a factory. The last three give a feel for the range of the noise – the ‘average’ level, the low background level and the highest level heard over a period of time. They are measured by taking a continuous recording of the dBA level over a known time period and then performing a statistical analysis of the record.

Two other measures are also commonly used.

*Day-night equivalent level ( $L_{dn}$ ).* This is a 24 h  $L_{eq}$  used for assessing community noise exposure, where the value of  $L_{eq}$  is measured over 24 h but the night-time readings between 22.00 and 6.00 h are weighted with the addition of 10 dB.

*Sound exposure level (SEL or  $L_{AE}$ ).* This is useful for comparing unrelated noise events in terms of their total acoustic energies: the energy is integrated over the duration of each event and expressed as the equivalent level over 1 s. It is also often used to describe the noise energy of a single event such as a passing car.

## FREQUENCY ANALYSIS

Frequency analysis is used when detailed information is needed about a complex sound signal. It can help to isolate possible sources of noise in machinery or in the wider environment. It is also used to evaluate the relative contributions of different frequency components when assessing the risk of damage to hearing. The most common analysis is to plot a frequency or power spectrum, which displays the sound energy across the range of frequencies. The frequency range can be split into frequency bands (usually one octave or 1/3 octave wide) which are analysed separately.

An octave is an interval which represents a doubling in frequency: the upper frequency bound is twice the lower frequency bound and the centre frequency is taken as the geometric mean of the two bounds. The octave bandwidths used for industrial sound analysis are normally



similar to those forming the musical scale, but the two sets of octave bands are not identical. The centre frequencies of the two sets are presented as follows:

#### Musical Octaves

32 64 128 256 512 1024 2048 4096 Hz

#### Industrial Sound Octaves

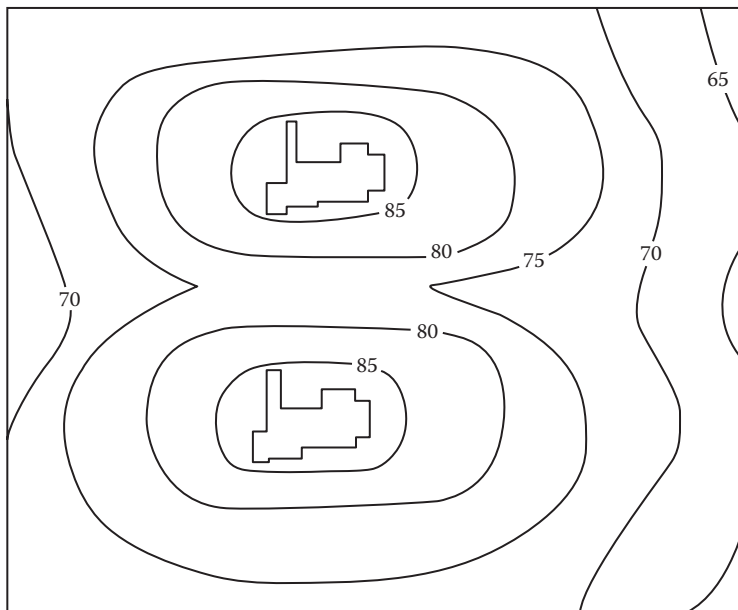
37.5 75 150 300 600 1,200 2,400 4,800 9,600 19,200 Hz

When more accurate frequency analysis is required, 1/3 octave bands can be used by logarithmically dividing each octave band into three. Narrowband analysis (using a spectrum analyser) gives still more detail. The bandwidth is usually defined as a fixed percentage of the frequency to be analysed. For example, a 6% frequency band at centre frequency 500 Hz would have a bandwidth of 30 Hz covering the range between 485 and 515 Hz.

## NOISE MEASUREMENT PROCEDURES

Noise measurements can be taken in order to survey the levels in a working area (producing a noise map, as illustrated in Figure 26.3) and identify any hazards to hearing for people in that area, or to assess the noise level produced by a particular source.

General guidance on procedures is given in the following text, but standardised test procedures have also been developed to ensure accurate and comparable measurements for specific applications. ISO 3740 (ISO, 2000) is a guide to noise problems, covering the general procedures used in the measurement of noise and the evaluation of its effects on human beings. This document lists ISO standards which are applicable to specific problems. In addition to international standards, national and industrial standards may apply.



**FIGURE 26.3** Noise map of a machine shop (with dB level contours).

## INSTRUMENTS

Sound level meters are the most commonly used instruments for measuring the auditory environment. Some incorporate filter sets giving octave band or 1/3 octave band frequency analysis. They record the sound level in each filter band to give the frequency content of the sound. In specialist applications, wave analysers may be used for more detailed narrowband frequency analysis. Integrating sound level meters are capable of measuring over longer periods of time and of providing statistical distribution analysis of the temporal characteristics of the sound levels.

Dosimeters are small integrating sound level meters which can be carried in a pocket or attached close to the wearer's ear and are used to measure the total personal exposure to noise during a period such as a working day. These give the total A-weighted sound energy received during the measurement period and may be used to calculate the  $L_{eq}$  value. The dose is often displayed as the proportion of the maximum permitted 8 h dose, and there may also be an indication of whether a standard peak noise level has been exceeded.

Probe microphones are used for measuring sound in the ear, for instance for evaluating the effect of hearing protectors. An alternative method of measuring the sound in earphones is to use an acoustic coupler (artificial ear), which is a standard cavity having an acoustic impedance similar to that of the ear.

## FIELD MEASUREMENTS

Noise from individual machines, tools or other equipment may be measured in the laboratory or in an anechoic chamber if precise measurements are required. For most purposes, however, this is not necessary and it is sufficient to use a convenient period in the workplace itself. Environmental measurements of course have to be carried out in the field and under normal conditions.

There are obviously problems in measuring individual noise sources in a workplace or in a town environment, due to the presence of other noise sources. It may be possible to choose a quiet period at night, or at the weekend, when very little other machinery is working, but it is unusual to have no background sources of noise. Even at night, there is likely to be traffic noise and, in the built environment, there will be heating, air conditioning, refrigeration or other plant running. However, reference to Table 26.3 shows that this will usually have a small or negligible effect on the noise measurement. If the difference between noise source and background is less than 3 dB (in any frequency band), or if the noise level is below that of the background, the noise level cannot be measured reliably. If the difference is between 3 and 10 dB, the effect of the noise source can usually be calculated to an acceptable degree of accuracy by comparison with the background level measured with the noise source switched off. This is more difficult when background noises are intermittent or fluctuating, and in any field survey care has to be taken to monitor this during the measurement period.

Outdoors several factors may affect the measurements. Wind is the most common problem, since this causes air turbulence around the microphone and low frequency noise. It is usual to shield the microphone with open-cell foam when taking measurements outside, although this is not fully effective at higher wind speeds. Wind, temperature and humidity can also affect the actual sound levels, as they change attenuation over distance and can create shadow zones. A fuller discussion of these atmospheric effects, and the attenuation produced by surroundings and barriers or walls, can be found in Beranek (1971).

## LABORATORY MEASUREMENTS

Although background noise is likely to be less of a problem in the laboratory, measurements there are liable to be affected by the structure of the enclosed space (by walls or other objects in the acoustic field). Sound may be reflected off hard surfaces or alternatively may be absorbed by fabric surfaces or acoustic tiles. When sound is reflected, the distribution of sound levels in the room will depend on the

phase relationships between the incident and reflected sound waves. This may increase noise levels or interfere with speech and other signals. Sound can be reflected several times from the walls of a room and cause reverberation. (Reverberation time is the time taken for the resultant sound to decay.)

In enclosed spaces, there are three distinct regions around a noise source:

*Near field*, where there may be interference between the sound wave being emitted and reflections from the room, so that the sound level varies slightly with the location of the meter. The near field is considered to extend over a distance approximately equal to the wavelength of the lowest frequency emitted or to twice the greatest dimension of the source machine (whichever is the greater).

*Far field*, which approximates to free-field conditions. Here the noise intensity decreases with distance from the source, following an inverse square law. This region can be identified by checking whether the sound level measurements obey this law.

*Reverberant field* close to reflecting surfaces, where the reflected noise is diffuse and noise levels are higher than in the free field and depend on the geometry of the room and the absorption properties of the surfaces. Measurements of the noise levels from the source will not be accurate in this region.

Two specialist types of noise laboratories are found: the anechoic chamber and the reverberation chamber. In an anechoic chamber, all surfaces are covered with highly sound absorbent material so that there are no reflections or echoes. This simulates the free-field conditions experienced outdoors. A reverberant room is one in which the walls completely reflect sound energy and where no walls or surfaces are parallel. This gives a diffuse sound field in which sound energy is uniformly distributed. This is most suitable for measuring the total power output of a noise source. Both the anechoic chamber and the reverberation chamber can be used to determine the sound power of a source, but an anechoic chamber can also be used to determine the reflectivity of surfaces.

Most laboratory measurements are in fact made in a semi-reverberant room, where there is a mixture of direct and reflected sound. To ensure accuracy, measurements of noise levels from a particular source should be made in the far-field region.

## LOCATION OF MICROPHONE

The person taking noise measurements and the microphone of the sound level meter itself can both interfere with the acoustic field being measured by blocking or reflecting sound waves, and this can cause significant errors (as large as 6 dB) in the measurements. If possible, the sound level meter should be left mounted on a stand. If it has to be held in the hand, it should be held at arm's length.

In general, whether outdoors or in an anechoic chamber, a directional (free field or frontal incidence) microphone is used and should be pointed towards the noise source. The alternative random-incidence microphones are designed to record sound from all directions and are normally used in reverberant or diffuse sound fields. However, it should be noted that some test standards prescribe procedures different from these. If a random-incidence microphone is used in free-field conditions, its readings are most accurate when it is orientated at an angle of 70°–80° to the source.

The microphone is normally located in a position representative of a hearer's ear but without any person present. If the hearer would be standing, their ear height is usually assumed to be 1.5 m. In a seated workplace, the seated ear height of an average person can be used for positioning the microphone. However, when assessing noise levels in a workplace, the measurements should be made at all positions in which people may be exposed to the noise, taking account of all tasks, operating conditions and working practices.

In the working environment, it is sometimes necessary for the machine operator to be present while the noise measurements are made. The microphone must then be positioned as close to the operator's head as possible, while avoiding reflections from the head or body or absorption by clothing. The microphone should be at least 50 mm from the side of the operator's head (and preferably further away). Where miniature instruments are attached to the hearer's collar or helmet, or placed inside hearing protectors, it is necessary to take account of such factors and to make corrections to the measured values (following the instructions which would normally be included in the instrument manuals).

## IDENTIFYING NOISE SOURCES

Various methods can be used to identify or isolate specific sources of noise in a complex environment. Where possible, the different components of the machine or environmental noise should be switched on or off separately and investigated in turn. If necessary, highly directional probe microphones can be used to determine the direction of sound propagation and thus identify the noise source, but simpler methods can also be used. For example, potential sources can be isolated by surrounding them with a suitable absorbent material, such as lead sheet, which can easily be shaped around components. Alternatively, they can be shielded by barriers or temporary walls of attenuating materials.

In isolating noise sources, it may be necessary to consider whether reflection or reverberation is contributing to the noise level. In some cases, housing and panels may act as a sounding box, either through vibration or by reflecting noise from other sources. If so, it may help in the investigation to fill the air spaces and damp the reverberations temporarily. Weights may be added to increase damping while testing the effects of different components.

## MEASUREMENT OF HEARING ABILITY AND RISK OF HEARING DAMAGE

### MEASURING HEARING ABILITY

A person's hearing ability is measured in terms of the minimum threshold of perception, using the normal psychophysical techniques such as the 'method of limits'. The measurements are usually made with an instrument called an audiometer and presented in the form of an audiogram. An audiometer generates pure tones over the range of audible frequencies and, when measuring hearing, scans through the range of intensities at each frequency. The tones are presented through headphones, and the signal can be directed separately or simultaneously to the two ears. If sounds are presented through speakers instead of headphones, there may be some differences in the measured threshold (perhaps up to 6 dB lower) due to differences in the acoustic field and the binaural threshold is likely to be lower than the monaural threshold (Loeb, 1986).

In order to avoid any influence from other sounds, no words should be spoken and the person whose hearing is being tested should be asked to indicate the presence or absence of a noise by a silent signal such as the movement of a finger. The measurements should obviously take place in a very quiet room. An initial period should be allowed for the person to adapt to the low noise level and recover from any temporary threshold shift (TTS) due to previous exposure to noise. The US regulations on occupational noise exposure (OSHA, 1983) specify that the first baseline audiograms taken in a hearing conservation programme should be taken after at least 14 h without exposure to workplace noise (unless wearing hearing protectors), although subsequent annual audiograms are permitted at any time during the working day.

'Normal' thresholds have been established (given, for example, in ISO, 1990a), and loss or impairment of hearing is usually defined as a minimum change of 10 or 15 dB in the threshold. Hearing level is taken as the amount by which the average threshold is raised in an individual, so that a positive hearing level represents hearing ability worse than the defined 'normal' level at any frequency. In an audiogram, the reference level of 0 dB represents the normal threshold of hearing (that of people who have no hearing disability or age deterioration) at each frequency over the auditory range.

### MEASURING MASKING THRESHOLDS

Although absolute hearing threshold is measured in a quiet environment, it is often important to know the hearing threshold for a signal or for speech in a noisy environment. This is termed the masking threshold, which defines the threshold of detection (not intelligibility) against the background level of noise. The effect is measured by determining the absolute threshold of the sound

when presented alone, then measuring it in the presence of the masking sound. The amount of masking is then defined as the difference in decibel level by which the threshold of audibility is raised above the absolute threshold.

The probability that a signal will be detected increases with the level of the signal above the background. So the masked threshold is sometimes defined as the level at which there is a given probability (say 75%) of correct detection of the signal. This is not an absolute measure since it depends on factors such as the level of expectancy of the hearer and the rise time, duration and temporal shape of the signal (Sorkin, 1987).

## ASSESSING THE RISK OF HEARING DAMAGE

Although a single impulsive sound can cause damage to the ear drum, this is rare and hearing loss is usually caused by long-term exposure to noise. The exposure does not have to be continuous to cause damage, since the effects of intermittent exposure are cumulative. Evidence of the risk of hearing damage was presented in ISO standard 1999 (ISO, 1990a), which clearly showed that repeated exposure to noise of over 90 dB for 8 h a day will produce significant deafness and that the potential for damage increases with the frequency of the noise source. It is now generally recognised that a risk of hearing damage exists at lower levels of noise, and most countries have established regulations for a maximum acceptable level for workplaces based on a time-weighted average sound level over an 8 h working day (noise dose). Usually three action levels are defined, at which employers must introduce progressively stricter administrative and engineering controls to protect their employees. For example in the European Union, the first two action levels are at 80 and 85 dBA, with in addition of an upper limit value of 87 dBA (Commission of the European Union, 2003). Employers are required to assess risk to workers' health and provide them with information and training if the noise level reaches 80 dBA (daily or weekly average), and to provide hearing protection and define hearing protection zones if the noise level reaches 85 dBA. There is a maximum workplace exposure limit of 87 dBA, taking account of any reduction in exposure provided by hearing protection that is worn.

Some standards also set an upper limit for peak SPL. In the European Union, an upper limit of 140 dBC (relative to a 20  $\mu$ Pa baseline) is set for impulsive noise exposure, with two lower action levels of 135 and 137 dBC (Commission of the European Union, 2003). In America, exposure to impulsive or impact noise should not exceed 140 dB peak SPL (OSHA, 1983). However, there is not a consensus on standards for exposure to infrasonic or ultrasonic frequencies. The risks from the complex effects of impulsive noises are reviewed in both Loeb (1986) and Kryter (1985), who quote damage risk contours, giving an indication of acceptable maximum peak SPL values for impulsive noises of different durations and repetition frequencies.

The main indicator used to assess hearing damage for an individual is their hearing threshold. If this is measured before and after exposure to loud noise, the threshold shift may be determined by the difference between the two thresholds.

TTS is a measure of the short-term and reversible change experienced after exposure to loud noise, which may persist for minutes or hours depending on the exposure. Since recovery starts as soon as the noise ceases, it is necessary to specify the time at which the TTS is determined. This is normally measured 2 min after exposure.

*Noise-induced permanent threshold shift* (NIPTS) is a measure of the long-term effect of exposure to noise, which is not reversible.

The mechanisms of TTS and NIPTS are not necessarily identical, but there seems to be a correlation between them, and Kryter (1985) has suggested that the TTS of a group of workers after 8 h exposure might be used as a criterion for the risk of long-term damage. More usually, hearing loss is assessed by the drop below the population norm. In industrial hearing conservation programmes, the operators at risk are assessed at the start of their employment and changes in hearing are monitored by annual tests. Kryter (1994) provides an extensive review of the evidence for the relative

influences on hearing thresholds of presbycusis, sociocusis and nosocusis (physiological ageing, exposure through activities of everyday living, and damage through disease or trauma, respectively).

Since industrial workers are exposed to noise which may vary considerably over a working period, it is important when assessing the risk of hearing damage to calculate the noise levels over the whole period of exposure (usually over an 8 h working day). The noise dose is calculated from the measured  $L_{eq}$  value (equivalent level of sustained noise). In measuring this, it should be remembered that operating conditions, especially communication signals, may contribute to the total noise dose. When this occurs in occupations where the signals are presented through headphones (as for aircraft pilots), the noise levels can be recorded at the ear, if necessary using a miniature or probe microphone, and the noise dose can be obtained by later analysis of the recording (Glen, 1976).

According to Davies and Jones (1982), most standards apply the 'equal-energy principle', which suggests that exposure to higher intensities can be permitted for short periods, provided that the total energy within an 8 h period does not exceed the normally permitted dose (as in ISO 1999 [ISO, 1990a]). This allows for quiet periods during the day, but does not take into account any recovery which may occur during these periods. For example, the equal energy principle assumes that a continuous 4 h exposure followed by 4 h in a quiet environment has the same effects as four 1 h exposures to the same level of noise separated by quiet periods of 1 h. On this assumption, halving the duration of exposure permits the sound level to be increased by 3 dB. The OSHA standard 1910.95 (OSHA, 1983) uses a less conservative 5 dB halving rule. However, to date, the extent of recovery during quiet periods is not fully understood, so that guidelines cannot yet be given for exposure/rest cycles.

## MEASUREMENT OF EFFECTS OF NOISE ON COMFORT AND PERFORMANCE

International and industry-wide standards have been drawn up for assessing comfort and the effects of noise on performance in different auditory environments. The effects of noise on performance are task specific and must often be assessed by experimentation, although general guidelines can be found in most ergonomics textbooks. Techniques for assessing two important aspects of communication (speech interference and design of auditory signals) are discussed later.

People vary enormously in their responses to noise, being influenced by factors such as intrusion into privacy or whether the sound is intermittent or unexpected, and probabilistic or statistical measures are needed to make any assessment of annoyance and performance effects of noise. Some of the reasons for individual differences are discussed by Jones and Davies (1984). People can adapt well to a continuous background of sound, and experiments therefore have to be designed carefully to include an adequate degree of realism in the experimental conditions. A corollary to this is that measurements of the level of annoyance may be influenced by factors other than the noise levels which are the subject of the investigation: noise annoyance may just be a symptom of poor morale or stress.

## ANNOYANCE

Many different measures are used to assess noisiness, annoyance and intrusiveness, some specific to environmental, community or transport noise. It is important to realise that loudness or noisiness cannot be equated with annoyance and investigators have found large differences in assessments of noise sources judged on these criteria (Kryter, 1985). Various aspects of perceptions of noisiness have been studied: judgements of noisiness or loudness of the neighbourhood, dissatisfaction with present noise level, frequency with which annoyance is felt, degree of annoyance and interference with activities (Jones and Davies, 1984). These should be carefully distinguished in any survey which is conducted, since measures taken simply to reduce noise levels may not be appropriate to reduce the level of annoyance. Context is also important. For example, an overheard conversation would be more annoying in a library than in an airport – partly due to expectation, partly due to



the lower degree of masking from background noise and partly due to the task being performed. Noise annoyance is obviously a multi-factorial problem. For example, at least 13 primary and derived measures have been used to assess community noise (Sanders and McCormick, 1992). Just a few of the evaluation methods which can be used are mentioned in the following text.

The annoyance level of noise is frequently measured by means of surveys and questionnaires, but it must be remembered that the wording of questions or of instructions given to the respondents can have a considerable influence on their responses and ratings. Advice on the design of appropriate scales for rating noise may be found in Loeb (1986) (and elsewhere in this book in more general terms).

Kryter (1985) developed scales of annoyance or perceived noisiness (in units called noys) which have been used for measures of aircraft and traffic noise. These were based on experimentally derived equal annoyance curves (pN dB) analogous to the equal loudness curves (phons). However, annoyance has additional psychological, social and economic dimensions so that Kryter noted that the threshold of perceived annoyance can vary with previous exposure, location indoors or outdoors, time of day and the impulsive or startle characteristics of the noise.

Many indices have been used to quantify disturbance caused by fluctuating noise (e.g. noise rating [NR], noise and number index [NNI], traffic noise index [TNI] and noise pollution index [NPI]), and these are discussed in Loeb (1986) and Kryter (1985). The values of these indices are often calculated by statistical analyses of noise over long duration measurement periods.

Schultz (1978) produced a dosage response curve as a predictive tool. This relates noise exposure (in terms of day-night average sound level  $L_{dn}$  in dBA) to the level of community annoyance, based on large-scale surveys from various countries (mostly related to traffic noise). Others suggest that short duration noises can be particularly annoying and have proposed measures using peak levels of noise events, as for aircraft overflights (Fidell, 1984). A discussion of some of these early measures can be found in Ollerhead (1973), while ISO 1996-1 defines measures and procedures for assessing various types of noise in community environments (ISO, 2003a).

Preferred noise criteria (PNC) curves were introduced as design criteria for background noise in offices, rooms or halls (Beranek, 1971). They were based on the speech interference level (SIL), which is a measure of speech intelligibility and is described in the next section. The PNC curves are a set of arbitrary sound spectra for steady noises, serving as a reference for rating noise environments. They are also useful in deciding where in the frequency spectrum the greatest benefits could be obtained by noise reduction treatments. PNC curves are based on the need for acceptable speech communication and specify that the loudness in phons should not exceed SIL by more than 22 units. The spectrum of the measured background noise is plotted over the noise criteria curves, and the noise is rated by the number of the curve which equals or just exceeds the noise spectrum at any point. This rating is then compared to a table of recommended ratings to evaluate the suitability of the noise level for the particular room environment.

It is not only loud noises that can be annoying. The effects of intrusiveness of low-level (or infrequent) noise exposure have been investigated by Fidell and others (Fidell et al., 1979; Fidell and Teffteller, 1981). They found that the  $L_{eq}$  measure is insensitive to these types of noises and developed methods of predicting the level of intrusion or annoyance from measures of signal detectability.

With the wide availability now of computer devices, sound meters can be used to provide real-time information in new, interesting ways, as, for example, in attempts to modify noisy or risky behaviour. Even schools can have devised noise-activated 'traffic lights' to show the children how noisy they are being at lunch time!

## SPEECH INTELLIGIBILITY

Noise can mask speech, but the degree of masking depends on various factors since there is a large measure of redundancy in speech, especially when the hearer is familiar with the subject matter. Thus, the assessment of speech intelligibility requires more than measuring the signal-to-noise ratio.

Speech intelligibility can be measured most directly by listening tests, presenting different types of standardised material, such as nonsense syllables, phonetically balanced word lists which contain all speech sounds (or phonemes) and sentences. The A-weighted decibel scale is normally used to measure sound level in these applications.

Several methods have also been developed to predict speech intelligibility from physical measures of the speech signal, noise environment and task parameters (e.g. speaker-hearer distance). Three of the methods are given here, and discussions of their relative merits can be found in Kryter (1994) and other textbooks. ISO standard 9921:2003 (ISO, 2003b) defines criteria for quality of speech communication, together with methods of measuring and predicting performance.

1. *Speech Intelligibility Index* (ANSI 1997; ISO 2003b) predicts the likelihood of difficulties with speech communication, and is calculated from the differences between the SPLs of the speech and the masking noise (the effective signal-to-noise ratio for the listener) in various frequency bands. These are weighted according to their importance in the intelligibility of speech and then summed to give the index.

Originally, for the 'articulation index', 20 bands were used between 250 and 7000 Hz, but an intelligibility index is now calculated from octave or 1/3 octave analysis. The relationship between the predictive index and intelligibility has been determined experimentally for different types of spoken material and expressed in a series of graphs which indicate the percentage of syllables, words or sentences that will be correctly understood. Corrections are made to take account of factors such as reverberation or noise interruption. The value of the speech intelligibility index indicates the extent of difficulty in speech communication (between 'difficulty likely' and 'good speech communication possible').

2. *SIL* (ANSI 1997; ISO 1974; ISO 2003b) is a simpler prediction method which does not require direct measurement of the speech level. The SIL is calculated from the mean of noise SPL in four octave bands. The resulting value is compared with tables relating it to speech intelligibility for different voice efforts (normal, raised, shouting) and for various distances between speaker and listener. ISO standard 9921 (ISO, 2003b) also gives a procedure for assessing SIL when the hearer is wearing hearing protectors.
3. *Direct Measurements* of SPL (in dBA) have also been used as an index of speech interference, predicting SIL. Loeb (1986) suggested a relationship of  $SIL = SPL - (9 \text{ or } 10 \text{ dB})$ .

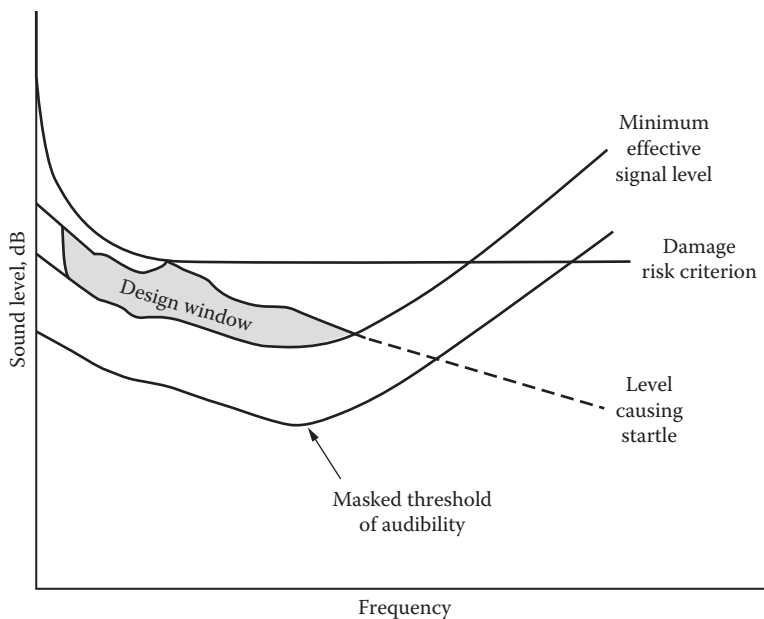
Webster (1979) produced a chart relating dBA, voice effort and speaker-hearer distance to quality of communication, so that either SPL or SIL could be used in assessing this.

## DESIGN OF WARNING AND INFORMATION SIGNALS FOR SAFETY AND AUDIBILITY

The effects of masking on warning and alarm signals are discussed in Webster (1984), giving a method for predicting the level at which a pure tone signal will be audible. However, additional criteria are needed to ensure that the signal also attracts attention and is recognisable. The perception of a masked signal varies with frequency, signal-to-noise ratio and absolute level of background noise. (See Chapter 12 for further discussion of warnings in general.)

Four main factors need to be taken into account in designing auditory warnings and signals: audibility, noise dose, startle and discriminability (Coleman et al., 1984). Reports from Patterson and Milroy (1979) and Coleman et al. (1984) describe procedures which have been used to design the appropriate sound level for auditory warnings in aircraft and in the mining industry, respectively.

The first stage of this process is the measurement or prediction of the masked threshold as described by Coleman et al. (1984) and Coleman (1998). The general method employed to identify the 'design window' for a particular work environment (as shown in Figure 26.4) is to analyse a recording of the background noise and perform a detailed narrow band spectral analysis. This can then be used to calculate the masked threshold, incorporating a 97.5th percentile absolute



**FIGURE 26.4** ‘Design window’ method for specifying signals to be used in noisy environments. (After Coleman, G.J. et al., *Communications in Noisy Environments. Technical Memorandum TM/84/1 (EUR P.74)*, Institute of Occupational Medicine, Edinburgh, U.K., 1984.)

threshold criterion. This ‘design window’ defines the upper and lower boundaries of frequency and intensity within which the signal components must lie in order both to be audible and to attract attention without causing startle or risk of hearing damage.

Figure 26.4 shows how the various factors are considered in specifying the ‘design window’ for a warning signal. This is constructed from knowledge of the hearing abilities of the workforce, the nature of the masked threshold due to the background noise and the levels above threshold which are necessary for the signal to be audible without causing startle. It would also be possible to take into account the attenuation effects when hearing protectors are used. The threshold could also be set for different population criteria: the maximum level to avoid startle should be determined for people with normal hearing, while the minimum effective level should be set for the people in the population who have the poorest hearing ability.

## ASSESSMENT OF PERSONAL HEARING PROTECTORS

When hearing protection is necessary in a noisy environment, criteria used in the selection of appropriate equipment must include the effect on communication (both for speech and for auditory signals) as well as the level of protection provided. New forms of both active and passive hearing protection devices are being developed, and these may require new assessment criteria (including assessment of usability in relation to adjustable tuning and attenuation). Behavioural issues also need to be considered, and Hughson et al. (2002) have suggested ways in which workers exposed to noise can be encouraged to wear hearing protection.

## EVALUATING ATTENUATION OF NOISE LEVELS

Noise attenuation curves are normally measured by the manufacturers and are supplied with hearing protection equipment. The attenuation may be measured by methods such as those in ANSI standard S12.6-2008 (ANSI, 2008). Manufacturers test hearing protectors to this standard and label

them with a noise reduction rating (NRR). The NRR is calculated from 1/3 octave band analysis of measured attenuation. This can then be used to estimate the noise exposure (in dBA) for the wearer: either as the workplace sound level in dB<sub>C</sub> less NRR, or as the sound level in dBA plus 7 dB less NRR (OSHA, 1983).

Sutton and Robinson (1981) reviewed various other procedures which can be used to estimate the level of protection given to a wearer (usually defined as the reduction in dBA level at the ear) from a knowledge of the frequency response of the protector and the frequency spectrum of background noise. The most accurate method of calculating the protection provided is given in ISO 1999 (ISO, 1990a), where attenuation (minus a variance correction) is subtracted from the workplace sound levels measured with octave band analysis. It should be noted, however, that these methods indicate the protection provided under optimum conditions and do not allow for poor fitting of the protector or noise leakage due to the wearer's hair or other factors. Studies have indicated that the methods may well overestimate the actual protection when measured in workplace conditions (Berger, 1983; Lempert, 1984).

A subjective method of measuring attenuation (by threshold shift) can also be used, as outlined in ISO 4869-1 (ISO, 1990b). A minimum of 16 people should be tested (preferably more), using normal audiometric techniques. When using the hearing protection, they should have a hearing threshold level in either ear which is not worse than 15 dB at frequencies below 2 kHz and not worse than 25 dB at frequencies above 2 kHz. The test signals specified are pink-noise-filtered through 1/3 octave bands from 63 to 8000 Hz. (Pink noise contains the full range of audible frequencies but is weighted towards the lower frequencies so that the sound pressure spectral density is inversely proportional to frequency.) This method can be used to compare or rank protection equipment and to evaluate design features which may affect performance.

The Health and Safety Executive (2005) guidance on control of noise regulations explains how to calculate the 'assumed protection' from the attenuation values provided by equipment manufacturers or recorded in tests on a group of people. The 'assumed protection' is taken as the mean attenuation at each frequency minus the standard deviation over the tests (probably of the order of 5 dB), in order to allow for the variation in protection between wearers and to ensure that the majority of wearers will have at least the predicted noise reduction.

Simulated ('dummy') head forms and miniature microphones can also be used to measure attenuation provided by hearing protectors, although the results may differ from subjective measurements because of the effects of bone conduction leakage. A semi-objective technique, in which a miniature microphone is attached to the wearer's ear, can be used to investigate the effects of hair, spectacles or helmets worn at the same time as hearing protectors.

## EVALUATING AUDIBILITY OF COMMUNICATIONS

In most work situations, the effects on speech communication need to be considered when assessing the effectiveness of hearing protectors. Evaluation measures include the speech intelligibility index, SIL and experimental testing of the intelligibility of messages (with subjective assessments). Although the criteria defined for intelligibility usually assume that the hearers have normal hearing ability, a measure of 'articulation index incorporating hearing ability (AIIHA)' was suggested by Coleman et al. (1984). They also proposed a technique for selecting protectors which would take into account the interaction between the attenuation characteristics and hearing ability, the range of noise and the speech conditions.

## CONCLUDING REMARKS

It is apparent from the wide variety of techniques introduced in this chapter that one of the most important considerations in any noise assessment is the choice of appropriate criteria for the evaluation. It is rarely sufficient simply to measure the SPL or acoustic power at a single location, since

the effects of noise on hearers are considerably more complex. Both perception and individual responses or preferences are usually influenced by the social or occupational context in which the sound is heard as well as by the auditory characteristics of the human ear.

When assessing a noisy environment, ergonomists may need to consider the effects on auditory health, safety, performance, transmission of information/communication, annoyance and pleasure for different groups of hearers. These are evaluated in different ways, and a range of criteria may be needed to assess the environment. It is hoped that this chapter gives some guidance on the most appropriate criteria and methods to choose for assessments of community, transport and industrial environments.

## REFERENCES

- ANSI (1997). *American National Standard Methods for Calculation of the Speech Intelligibility Index*. ANSI/ASA S3.5-1997 (R2012) (New York: American National Standards Institute).
- ANSI (2008). *Methods for Measuring the Real-Ear Attenuation of Hearing Protectors*. ANSI/ASA S12.6-2008 (New York: American National Standards Institute).
- Beranek, L.L. (1971). *Noise and Vibration Control* (New York: McGraw-Hill).
- Berger, E. (1983). Using NNR to estimate the real world performance of hearing protectors. *Sound and Vibration*, 18(5), 26–39.
- Coleman, G.J. (1998). The signal design window revisited. *International Journal of Industrial Ergonomics*, 22(4–5), 13–318.
- Coleman, G.J., Graves, R.J., Collier, S.G., Golding, D., Nicholl, A.G.McK., Simpson, G.C., Sweetland, K.F. and Talbot, C.F. (1984). *Communications in Noisy Environments*. Technical Memorandum TM/84/1 (EUR P.74) (Edinburgh, U.K.: Institute of Occupational Medicine).
- Commission of the European Union (2003). *Council Directive of 6 February 2003 on the Minimum Health and Safety Requirements Regarding Exposure of Workers to the Risks Arising from Physical Agents (Noise)*. Council Directive 2003/10/EC. *Official Journal of the European Union*, No. L42/38, 15 February 2003.
- Davies, D.R. and Jones, D.M. (1982). Hearing and noise. In: *The Body at Work*, W.T. Singleton (ed.) (Cambridge, U.K.: Cambridge University Press), pp. 365–413.
- Fidell, S. (1984). Community response to noise. In: *Noise and Society*, D.M. Jones and A.J. Chapman (eds.) (Chichester, U.K.: John Wiley & Sons), pp. 247–277.
- Fidell, S. and Teffteller, S.R. (1981). Scaling the annoyance of intrusive sounds. *Journal of Sound and Vibration*, 78, 291–298.
- Fidell, S., Teffteller, S.R., Horonjeff, R.D. and Green, D.M. (1979). Predicting annoyance from detectability of low-level sounds. *Journal of the Acoustical Society of America*, 66, 1427–1434.
- Glen, M.C. (1976). The contribution of communications signals to noise exposure. *Applied Ergonomics*, 7(4), 197–200.
- Health and Safety Executive (1995). *Sound Solutions – Techniques to Reduce Noise at Work* (Sudbury, Ontario, Canada: HSE Books).
- Health and Safety Executive (2005). *Controlling Noise at Work. Guidance on the Control of Noise at Work Regulations 2005*. Guidance on Regulations L108 (Sudbury, Ontario, Canada: HSE Books).
- Hughson, G.W., Mulholland, R.E. and Cowie, H.A. (2002). *Behavioural Studies of People's Attitudes to Wearing Hearing Protection and How These Might Be Changed*. Health and Safety Executive Research Report No. 028 (Sudbury, Ontario, Canada: HSE Books).
- ISO (1974). *Acoustics – Assessment of Noise with Respect to Its Effect on the Intelligibility of Speech*. ISO Technical Report TR 3352 (Geneva, Switzerland: International Standards Organisation).
- ISO (1990a). *Acoustics – Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment*. ISO 1999 (Geneva, Switzerland: International Standards Organisation).
- ISO (1990b). *Acoustics – Hearing Protectors – Part 1: Subjective Method for the Measurement of Sound Attenuation*. ISO 4869-1 (Geneva, Switzerland: International Standards Organisation).
- ISO (2000). *Acoustics – Determination of Sound Power Levels of Noise Sources – Guidelines for the Use of Basic Standards*. ISO 3740 (Geneva, Switzerland: International Standards Organisation).
- ISO (2003a). *Acoustics – Description, Measurement and Assessment of Environmental Noise – Part 1: Basic Quantities and Assessment Procedures*. ISO 1996-1 (Geneva, Switzerland: International Standards Organisation).
- ISO (2003b). *Ergonomics – Assessment of Speech Communication*. ISO 9921:2003 (Geneva, Switzerland: International Standards Organisation).

- Jones, D.M. and Chapman, A.J. (1984). *Noise and Society* (Chichester, U.K.: John Wiley & Sons).
- Jones, D.M. and Davies, D.R. (1984). Individual and group differences in the response to noise. In *Noise and Society*, D.M. Jones and A.J. Chapman (eds.) (Chichester, U.K.: John Wiley & Sons), pp. 125–153.
- Kohler, H.K. (1984). The description and measurement of sound. In *Noise and Society*, D.M. Jones and A.J. Chapman (eds.) (Chichester, U.K.: John Wiley & Sons), pp. 35–76.
- Kryter, K.D. (1985). *The Effects of Noise on Man*, 2nd edn. (London, U.K.: Academic Press).
- Kryter, K.D. (1994). *The Handbook of Hearing and the Effects of Noise* (San Diego, CA: Academic Press).
- Lempert, B.L. (1984). Compendium of hearing protection devices. *Sound and Vibration*, 18, 26–39.
- Loeb, M. (1986). *Noise and Human Efficiency* (Chichester, U.K.: John Wiley & Sons).
- Ollerhead, J.B. (1973). Noise: How can the nuisance be controlled? *Applied Ergonomics*, 4(3), 130–138.
- OSHA (Occupational Safety and Health Administration) (1983). Occupational noise exposure; hearing conservation amendment; final rule. *Federal Register*, 48, 9738–9783.
- Patterson, R.D. and Milroy, R. (1979). *Existing and Recommended Levels for Auditory Warnings on Civil Aircraft*. Civil Aviation Authority Contract Report (Contract No. 7D/S/0142) (Cambridge, U.K.: Medical Research Council, Applied Psychology Unit).
- Sanders, M.S. and McCormick, E.J. (1992). *Human Factors in Engineering and Design*, 7th edn. (New York: McGraw-Hill).
- Schultz, J. (1978). Synthesis of social surveys on noise annoyance. *Journal of the Acoustical Society of America*, 64, 377–405.
- Sorkin, R.D. (1987). Design of auditory and tactile displays. In *Handbook of Human Factors*, G. Salvendy (ed.) (New York: John Wiley & Sons), pp. 549–576.
- Sutton, G.J. and Robinson, D.W. (1981). An appraisal of methods for estimating effectiveness of hearing protectors. *Journal of Sound and Vibration*, 77, 79–91.
- Webster, J.C. (1979). Effects of noise on speech. In *Handbook of Noise Control*, 2nd edn., C.M. Harris (ed.) (New York: McGraw-Hill).
- Webster, J.C. (1984). Noise and communication. In *Noise and Society*, D.M. Jones and A.J. Chapman (eds.) (Chichester, U.K.: John Wiley & Sons), pp. 185–220.



---

# 27 Anthropometry for Ergonomic Design

*David Feathers, Clive D'Souza and Victor Paquet*

## CONTENTS

|                                                                                         |     |
|-----------------------------------------------------------------------------------------|-----|
| Introduction.....                                                                       | 725 |
| Anthropometric Measurement Methods .....                                                | 726 |
| Body Measurements.....                                                                  | 726 |
| Functional Anthropometric Measurements.....                                             | 731 |
| Evaluation of Measurement Error.....                                                    | 733 |
| Applying Anthropometric Data in Ergonomic Design .....                                  | 733 |
| Determining the Target User Population.....                                             | 734 |
| Determining the Critical Design Dimensions and Anthropometric Dimensions .....          | 734 |
| Determining the Appropriate Source or Sources of Data.....                              | 735 |
| Determine the Percentage of the Target Users Population to Be Accommodated by Design... | 736 |
| Univariate Analyses to Inform Design .....                                              | 737 |
| Multivariate Analyses to Inform Design .....                                            | 737 |
| Necessary Adjustments for Enhanced Usability .....                                      | 738 |
| Other Considerations.....                                                               | 738 |
| Case Study: Evaluating Office Workstation Clearances for Wheelchair Users .....         | 739 |
| Target User Population.....                                                             | 739 |
| Critical Design and Anthropometric Dimensions .....                                     | 739 |
| Appropriate Source of Data .....                                                        | 739 |
| Percentage of the Target User Population to Be Accommodated by the Design .....         | 739 |
| Portion of the Distribution That Will Be Excluded and Accommodated .....                | 740 |
| Analysis.....                                                                           | 740 |
| Necessary Adjustments to Enhance Usability .....                                        | 744 |
| Conclusions.....                                                                        | 744 |
| References.....                                                                         | 744 |

## INTRODUCTION

Anthropometry is the study of human body sizes and abilities. Anthropometry is derived from the Greek words ‘anthropo’, meaning ‘human being’, and ‘metry’, meaning ‘measure’. Anthropometry provides descriptive information about the distributional characteristics of physical body dimensions and functional abilities of large samples of people. Physical anthropometric dimensions include basic whole-body measurements such as standing height, as well as specific body segment lengths, widths, depths, circumferences and weights. Functional abilities include things such as grip strength, push and pull strength, reaching capabilities, fields of vision and applied measurements specific to task performance.

Anthropometry has been applied as early as when ancient Greek and Egyptian artists used physical body measurements to make representations of people more realistic. Today, anthropometry is used in many disciplines. Physical anthropologists use anthropometry to investigate

human and non-human structure, function, abilities and comparative evolutionary changes. Clinicians use anthropometric information to determine when one's body size or functional characteristics fall outside statistical norms in order to aid diagnosis. Human modellers use anthropometry to ensure that digital representations of people are physically realistic. Biomechanists use anthropometry to estimate body segment centre of mass and joint centre of rotation characteristics needed to study forces and torques on the body. Standards makers consider anthropometric information when developing the physical requirements of public spaces, including building codes such as doorway widths, the heights of drinking fountains, the dimensions of stairways and many other building features. For engineers and designers, in particular, anthropometry provides information that can be used for the ergonomic design of workstations, furniture, equipment, tools and clothing. This chapter focuses on the application of anthropometry in engineering and design.

## ANTHROPOMETRIC MEASUREMENT METHODS

Standardised measurement methods exist to help ensure that anthropometric data are measured and reported in consistent ways. Recently, the International Standardisation Organisation (ISO) technical committee on ergonomics has developed ISO 7250: Basic Human Body Measurements for Technical Design (1996) to standardise language and measurement methods used in anthropometry, and ISO 15535: General Requirements for Establishing an Anthropometric Database (2003) to standardise the variables used and reporting methods of anthropometric studies. Textbooks such as the *Anthropometric Standardisation Reference Manual* (Lohman et al., 1988) and *Bodyspace: Anthropometry, Ergonomics, and the Design of Work* (Pheasant and Haslegrave, 2006, third edition) offer basic human body measurement techniques for landmark location, notes on the construction of dimensions and ergonomic design applications. These and other important sources of information listed at the end of this chapter are important for ergonomic and other human-centred design applications.

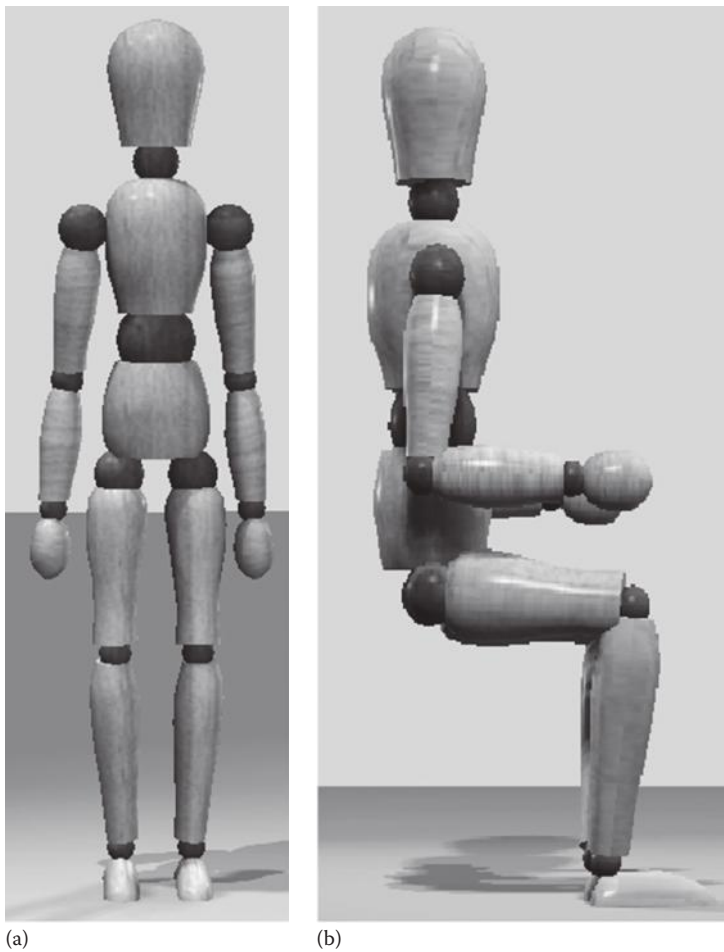
## BODY MEASUREMENTS

Body measurements usually require an individual to be measured while in a standardised standing or seated posture to help ensure measurement reliability (Figure 27.1). The traditional anthropometric standing position requires the individual to stand erect, with the ankles close together, arms relaxed to the side of the body, with the head positioned such that the corners of the eyes align horizontally with the ears. Standardised seated posture requires the individual to sit erect on a standard seat surface. The elbows and knees are placed and should remain at 90° flexion. The palms face medially (see Figure 27.1).

Other standardised postures have been developed more recently to allow the shapes of the body to be more accurately assessed by scanning technologies. A series of scans are assembled to build a complete 3-D view of that individual, including areas often difficult to view, such as the armpit (axillary) area. Figure 27.2 shows some of the postures that can be adopted to build a 3-D view of an individual for anthropometric measurement.

Observable or palpable body features, or 'body landmarks', are used to standardise the locations on the body from which measurements are made. Body dimensions include the vertical distance from a body landmark to the floor, the distances between anatomical landmarks or the circumferences of body segments. For example, the corner of one's eye (ectocanthus) may be used to measure standing or seated 'eye height' or the palpable bony protrusions of the right and left hip may be used to measure 'hip breadth'. Table 27.1 provides just a few of the hundreds of dimensions that have been reported.

Measurements can be made in a variety of ways. Simple devices used for the manual measurement of body dimensions have been used for many years. One of the most common traditional

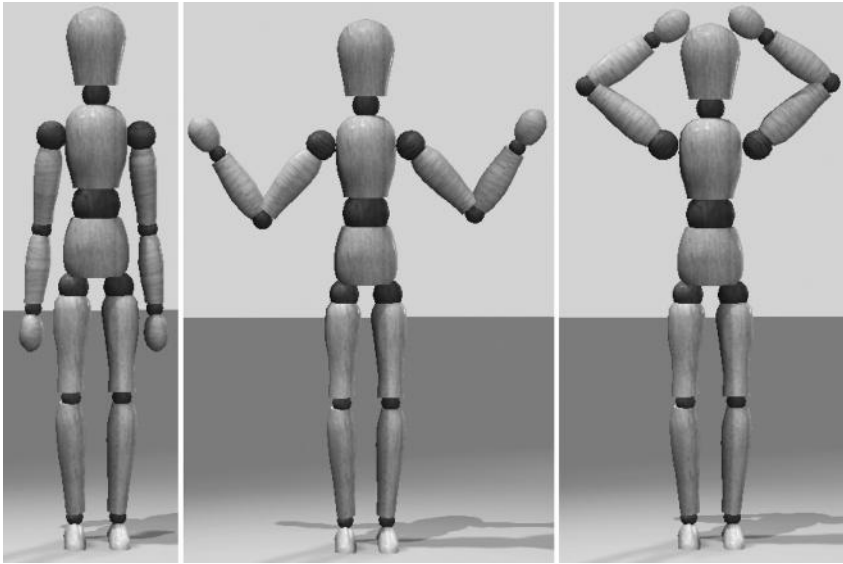


**FIGURE 27.1** Illustrations of the traditional (a) standing and (b) seated anthropometric postures.

instruments is the anthropometer, a tool having a rod and sliding perpendicular arm used to measure heights, widths and depths (Figure 27.3). A spreading calliper having two curved arms hinged together is used to measure body segment widths and breadths that may be on a complex surface (e.g. cranium). Graduated cones are used to measure grip circumferences, and tape measures are used to measure other circumferences such as the distance around the waist. A scale is used to measure body weight.

Photographs and other video methods have also been used to measure body dimensions in two dimensions. One such method, known as the Morant technique (Morant, 1947), uses grids that are attached perpendicularly and are placed behind and to the side of the person measured. A modern variant of this approach involves the use of digital video recording and requires manual digitisation of a reference measure to develop the appropriate scales based on the number of pixels between the ends of the reference. The anthropometric measurements are then obtained by manually digitising the body landmarks in the digital photograph, and the number of pixels between the landmarks to the number of pixels of the reference measure can be compared to obtain the measurement.

New technology-led methods of obtaining anthropometric data are also emerging. Contemporary computerised technology, such as laser scanning technologies and electromechanical approaches,



**FIGURE 27.2** Varied postures for scanning technologies to build a complete 3-D view.

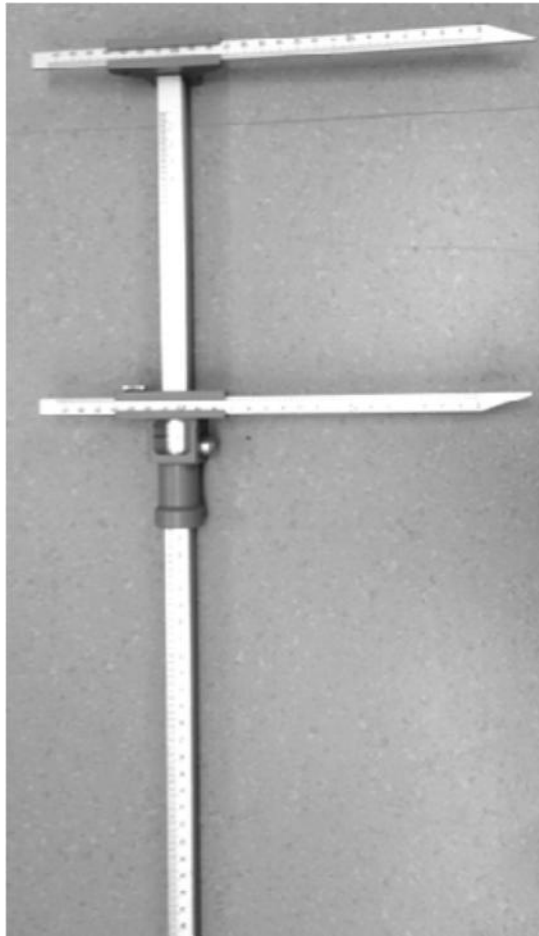
**TABLE 27.1**

**Some Examples of Traditional Anthropometric Dimensions**

| Heights                                  | Breadths                     | Depths                            | Lengths                                       | Circumferences                               |
|------------------------------------------|------------------------------|-----------------------------------|-----------------------------------------------|----------------------------------------------|
| Ankle height                             | Ankle breadth                | Abdominal depth, sitting          | Arm length, shoulder (acromion) to finger tip | Ankle circumference                          |
| Back of knee height (popliteus), sitting | Chest breadth                | Abdominal depth, standing         | Buttock to back of knee (popliteus), sitting  | Calf circumference                           |
| Ear height (tragion), standing           | Hand breadth including thumb | Ring finger, proximal joint depth | Buttock to front of knee, sitting             | Chest circumference at the armpits (axillae) |

is often used to collect information about body size and physical ability. The advantages of these systems include speed of data capture and excellent resolution of the body surface (morphological) output files in three dimensions. Other benefits are the ability to more directly develop 3-D human models, ability to evaluate multiple body dimensions simultaneously and improved ability to identify joint centres of rotation and movement for complex, real-life tasks such as reaching tasks within an automobile (Reed et al., 2003).

Laser-based technologies provide a way to measure distances on complex body surfaces. Most protocols require the placement of surface markers to ensure the proper location of bony protrusions that are used as measurement landmarks beneath the surface of the skin in order to standardise the measurements. Other protocols using this technology have morphological extraction algorithms that estimate these landmarks based on the body surface features. These methods have shown to have very good accuracy and repeatability (e.g. Brooke-Wavell et al., 1994). The Civilian American and European Surface Anthropometry Resource (CAESAR) project is one of the most notable recent anthropometric studies of North American and European civilians that used laser scanning methods (2400 North American and 2000 European subjects aged 18–65;



**FIGURE 27.3** The anthropometer has been used for many years to measure heights and widths of body segments.

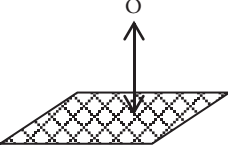
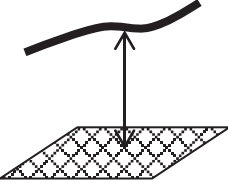
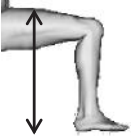
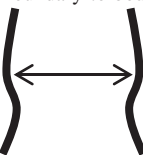
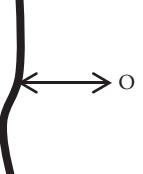
from SAE International). Other large anthropometric studies and reports have been performed in other countries such as Japan (e.g. Japanese Body Size Data, 1992–1994; with a sample size of 34,000 measured by the Research Institute of Human Engineering for Quality Life, 1997) and India (e.g. SIZE INDIA; with a sample size of 5000 measured by researchers at the Automotive Research Association of India, 2011).

Electromechanical or potentiometric systems can also be used for collection of digital anthropometric measurement collection. These devices allow the measurer to manually digitise points in 3-D space. The operator handles the instrument and guides the movements of a probe tip to manually to render discrete points or a surface scan of person's body area.

Three-dimensional anthropometric measurements do not have exactly the same formal anthropometric definitions as the corresponding 1-D or 2-D measurements. Traditional measurements are drawn from features that can be discerned by tools such as an anthropometer or calliper and are written to reflect 2-D tool use as well as 1-D or 2-D measurement procedure. Table 27.2 provides some examples of how measurement procedures may differ between 2-D measurements recorded with conventional instruments and 3-D measurements recorded with a 3-D electromechanical probe.

TABLE 27.2

### Measurement Procedures Will Vary by the Type of Measurement and the Types of Measurement Instruments Used

| Type of Measurement                                                                                         | Example                                                                                                                | Two-Dimensional Measurement Procedures                                                                                                                                                             | Three-Dimensional Measurement Procedures (e.g. Electromechanical Probe)                                                                                                            |
|-------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Body landmark-to-plane<br> | Acromion height to floor (seated)<br> | Palpate landmark, place anthropometric blade on landmark, ensure anthropometer is parallel to plane, record measurement                                                                            | Define in software program: plane (floor), landmark and dimensioning<br>Digitise: plane and landmark<br>Save program                                                               |
| Boundary-to-plane<br>     | Thigh height to floor<br>             | Locate boundary area, sweep anthropometer blade across boundary according to the definition of the measurement (i.e. superior-most portion in terms of the floor plane) and record the measurement | Define in software program: plane (floor), scan feature for the area to-be-scanned and dimensioning of the plane and scan results<br>Digitise: plane and scan area<br>Save program |
| Boundary-to-boundary<br> | Bi-deltoid breadth<br>              | Locate both boundaries, sweep anthropometer blades across both boundaries simultaneously<br>Record measurement                                                                                     | Define in software program: scan feature for boundary one and scan feature for boundary two<br>Digitise: scan both areas<br>Save program                                           |
| Boundary-to-landmark<br> | Buttock-knee length<br>             | Locate boundary, palpate for landmark, place on blade of the anthropometer on the landmark and sweep the other blade of the anthropometer across the boundary area<br>Record the measurement       | Define in software program: landmark and scan feature for boundary<br>Digitise: landmark and scan the defined boundary<br>Save program                                             |

(Continued)



TABLE 27.2 (Continued)

**Measurement Procedures Will Vary by the Type of Measurement and the Types of Measurement Instruments Used**

| Type of Measurement                               | Example                                                                                                 | Two-Dimensional Measurement Procedures                                                                                                                                      | Three-Dimensional Measurement Procedures (e.g. Electromechanical Probe)                                     |
|---------------------------------------------------|---------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| Landmark-to-landmark<br>O $\longleftrightarrow$ O | Bi-spinous breadth<br> | Palpate for body landmarks<br>Place the anthropometer blade on one landmark and extend the other blade of the anthropometer to the other landmark<br>Record the measurement | Define in software program: landmark one and landmark two<br>Digitise: landmark one and two<br>Save program |

## FUNCTIONAL ANTHROPOMETRIC MEASUREMENTS

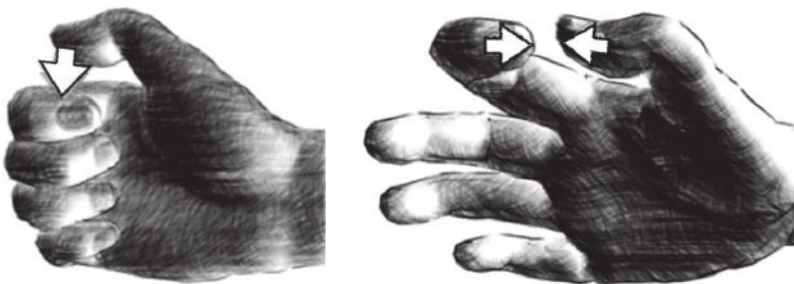
Functional performance measures often include standardised procedures for the measurements of isometric grip, push or pull strength and reaching capabilities. A hand dynamometer set at one or more grip circumferences is usually used to collect peak grip strength (see Figure 27.4) in a variety of grip positions (see Figure 27.5). Similarly, force gauges are used to measure isometric push or pull strength. Reaching abilities measured in a variety of ways including video-based methods, potentiometric methods and even methods that require an individual to mark maximum or comfortable reach locations with pen on a paper grid.

Functional hand anthropometry studies have often focused on able-bodied individuals stratified by gender and age (e.g. Boatright et al., 1997; Bohannon et al., 2006; Crosby et al., 1994; Hanten et al., 1999; Mathiowetz et al., 1985b) with an occasional emphasis on elderly populations (e.g. Bear-Lehman et al., 2003; Brennan et al., 2004; Desrosiers et al., 1995a; Fiebert et al., 1995; Horowitz et al., 1997; Rice et al., 1989; Shechtman et al., 2004), special populations such as persons using wheeled mobility devices (e.g. Floyd et al., 1966; Paquet and Feathers, 2004) and studies or publications providing these data in a design context (e.g. Damon et al., 1966; Imrhan and Loo, 1989a,b; Peebles and Norris, 1998; Steinfeld, 1986; Voorbij and Steenbekkers, 2002).

Most of these anthropometric studies that have measured peak isometric grip strength typically do so using a standardised set of instrumentation, grip formations and body positioning to either establish norms or facilitate comparisons with previously published normative data (Boatright et al., 1997; Bohannon et al., 2006; Brennan et al., 2004; Crosby et al., 1994; Desrosiers et al., 1995a; Hanten et al., 1999; Horowitz et al., 1997; Imrhan and Loo, 1989b; Mathiowetz et al., 1985b). Measurement of power grip strength usually employs the standardised arm posture as recommended by the American Society of Hand Therapists (Fess, 1992).



**FIGURE 27.4** Grip (left) and pinch (right) dynamometers for measuring upper extremity strength.



**FIGURE 27.5** Varied hand positions in preparation for dynamometric measurement.

Motion analysis systems provide measures of physical abilities related to the way individuals move (kinematics) and can be used along with other types of instrumentation, such as force plates or pressure transducers to provide biomechanical information or measures of balance. These systems allow positions of body landmarks to be tracked over time during a physical activity requiring movement. The data can be evaluated statistically or can serve as an example of a human task simulation. Such methods of data collection allow more lifelike dynamic digital human models that can be used to evaluate human performance in virtual environments (see Chapter 8).

## EVALUATION OF MEASUREMENT ERROR

Anthropometric measurements are subject to errors, which may stem from within the observer or between observers, the technology used, the treatment of the data and other sources. Research studies have described the nature, the sources and the impact errors have throughout the measurement process (e.g. Aldridge et al., 2005; Feathers et al., 2004; Kouchi and Mochimaru, 2011). Assessing the quality of the data collected is a responsibility of the anthropometrist. The technical error of measurement (TEM) has been used by anthropometrists to look into intra-observer error TEM and inter-observer TEM, where

$$\text{TEM} = \frac{\sqrt{\sum D^2}}{2N}$$

where

D is the difference between two measurements on the same individual  
N is the number of individuals measured (Lohman et al., 1988)

If more than two observers are involved, then the equation changes to

$$\text{TEM} = \sqrt{\sum^N [\sum^K M(n)^2 - (\sum^K M(n)^2/K)]/N(K-1)}$$

where

N is the number of subjects  
K is the number of measurements of the variable taken on each subject  
M(n) is the nth replicate of the measurement, where n varies from 1 to k

Reliability (R) is classically defined as

$$R = 1 - \left[ \frac{(\text{TEM})^2}{(\text{SD})^2} \right]$$

where SD is the total inter-subject variance, including measurement error (Lohman et al., 1988).

This coefficient of reliability reveals what proportion of the between-subject variance in a measured population is free from measurement error.

Consistent and reliable measurement may rely on information gathered from bottom-up (palpation, passive movement of the subject's limbs, looking for visual cues) and top-down (anatomical and anthropometric knowledge) information processing simultaneously. Modalities offering the bottom-up information may concurrently rely on visual cues of ink marks, stickers or bony protrusions and haptic palpation. Processing this bottom-up information under the framework of definitions given in a manual or knowledge accumulated through feedback gives the measurer the ability to arrive at an accurate measurement despite diverse body types measured.

## APPLYING ANTHROPOMETRIC DATA IN ERGONOMIC DESIGN

Anthropometric data should be collected for large samples of populations so that the distributional characteristics of one or more dimensions for a population can be estimated. There are many sources of anthropometric variability. No two individuals are exactly alike in terms of all body dimensions and physical abilities. An individual's body size also changes with

time, drastically during childhood and throughout the course of adulthood. As an individual approaches old age, for example, his or her height decreases and weight may fluctuate. There are also differences in the way dimensions may be distributed between diverse populations or different subsamples of populations. Differences often exist between men and women, different ethnicities, people of different geographic regions, age groups and occupations. The anthropometric dimensions of people have also changed systematically over time. In modern countries, individuals tend to be taller and heavier than those previous generations, perhaps because of the improved availability of food.

Anthropometric data of individual body size or functional dimensions are often summarised with descriptive statistics stratified by categories of gender and age, and sometimes stratified by ethnicity, geographic region of origin or other demographic variable. The variability of anthropometric dimensional values within such subgroups often takes the shape of a Gaussian distribution, the mean and standard deviation are used to describe the distributional characteristics of the dimension.

The *percentile value*, or the value of a dimension that is greater than or equal to a certain percentage of a distribution, can be used to evaluate the percentage of the population that a design may be accommodated or excluded in terms of a specific dimension. For example, the 95th percentile value for stature with a population represents the value of a dimension that is equal to or greater than 95% of the statures with that population. Contrary to the way many report anthropometric data, there is no such thing as a 95th percentile person; instead, there is only a 95th percentile value of a dimension. No one has exactly the same percentile value of all possible body sizes and functional abilities! For example, an individual may be average stature but have very large feet, say 95th percentile, which may require special ordering direct from a manufacturer!

Anthropometric data can be extremely valuable in the design of environments, equipment, tools and clothing. Use of anthropometric data in design requires determining

- The target user population
- The critical design and anthropometric dimensions
- The appropriate source of data
- The percentage of the target user population to be accommodated by the design
- The portion of the distribution that will be excluded, usually the largest and/or smallest values of the distribution
- The univariate or multivariate statistical analysis to inform design
- Necessary adjustments to enhance usability

## DETERMINING THE TARGET USER POPULATION

The first step in applying anthropometry to the design of an environment, product or equipment is to identify the users. Is there a specific user group for the product or are there a variety of user groups? Public environments will be used by children, adults, older adults and those with functional limitations, and therefore the spaces and design features should accommodate an extremely wide range of body sizes and abilities. In contrast, an industrial tool may be designed for a very specific user group or groups (e.g. male and female carpenters), having a narrower set of body sizes and abilities. Demographic data of the intended user groups can inform the design issues and can be used to determine whether there is an appropriate anthropometric database in existence or whether there is a need to collect new anthropometric data.

## DETERMINING THE CRITICAL DESIGN DIMENSIONS AND ANTHROPOMETRIC DIMENSIONS

It is important for designers to consider the relationships between the design dimensions (the design of the environment or product), the anthropometric dimensions (the body sizes and abilities) and

human performance requirements (e.g. how effectively must a person perform). For example, the relationship between handle dimensions, hand size, grip size and grip strength are important factors to consider when attempting to better fit hand-operated products and tools to reduce the risk of injury and improve productivity (Chaffin et al., 2006; Pheasant and Haslegrave, 2006). The relationship between the dimension and 'fit' must also be considered. Sometimes the relationship between the design specifications and the anthropometric dimensions requires designing to a narrow tolerance (e.g. tight fitting gloves and hand dimensions), while in other cases the relationship between the design specifications and anthropometric dimension is not as direct (e.g. diameter of a pen and hand dimensions). The basic postures and movements within the context of how the individual will interact with the design must also be considered. It is often the case that an anthropometric analysis will be applied to a combination of design dimensions simultaneously. Such approaches require multivariate statistical methods and/or digital human modelling approaches that are discussed later in this chapter.

### **DETERMINING THE APPROPRIATE SOURCE OR SOURCES OF DATA**

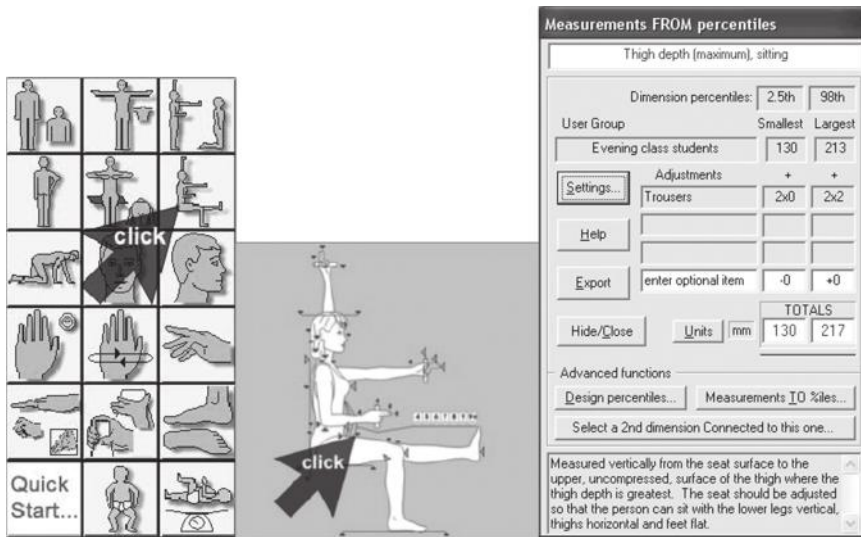
Once the user group(s), the critical design characteristic(s) and anthropometric dimension(s) are determined, the designer must then select the appropriate source or sources of data for the anthropometric analysis. Collecting anthropometric data requires a significant amount of anatomical knowledge and measurement expertise, and it can be very time consuming and costly. Therefore, ergonomics and human factors (E/HF) professionals and designers will usually rely on anthropometric databases of previous studies as opposed to collecting new anthropometric data.

Numerous data sources are available that describe the anthropometry of different populations, perhaps the most common being of healthy adult populations. Typically, the data are presented as percentile-based tabular representations (e.g. Anthropology Research Staff, 1978a, b; Gordon et al., 1989; Kennedy, 1986) or incorporated into software applications that feature percentile-based digital human models or manikins for computer-aided design (CAD) and ergonomics analysis. Some examples include Jack and Process Simulate Human (Siemens), Delmia SAFEWORK (Dassault Systemes), SAMMIE (SAMMIE Research Group, SAMMIE CAD Ltd.), HADRIAN (Marshall et al., 2002, 2010) and HumanCAD (NexGen Ergonomics, Inc.). The economic benefits and the compatibility of using digital human modelling tools with contemporary computer-aided design approaches have led to their widespread use in many different design domains (Chaffin, 2001).

The most comprehensive anthropometric studies have focused primarily of military personnel, with one of most comprehensive studies in the United States being the 1988 U.S. Army Anthropometric Survey (ANSUR) of that included 132 dimensions of approximately 9000 army personnel (Clauser et al., 1988). The attention to the anthropometry of military personnel is, at least in part, due to the critical need for the military to have information to provide well-designed uniforms, equipment, land vehicles and aircraft.

One of the most inclusive sources of civilian anthropometric data was a NASA technical report produced by the staff of the Anthropology Research Project in 1978. This report contains anthropometric data across a variety of civilian and military populations for a large number of anthropometric variables, including information about the mass distribution of body segments. Marras and Kim (1993) have provided a useful anthropometric data set of occupational populations, offering some preliminary anthropometric information about different U.S. occupational populations and demonstrating important differences in anthropometry between industrial and military populations.

More recently, the Civilian American and European Surface Anthropometry Resource (CAESAR) project used laser scanning to collect very detailed information about the body surface contours and sizes of approximately 2400 North American and 2000 European civilians from 1998 to 2000 (Robinette, 2000). Measurements were recorded with individuals in standing, standardised seated and relaxed seated postures. Thousands of anthropometric points on the body surface were collected with each scan, which allow detailed 3-D evaluations of surface areas and volumes rendered



**FIGURE 27.6** PeopleSize Software Interface (openerg.com, 2014).

and provide individual static human models that can be used to evaluate human–environmental fit in three dimensions. Since markers are also placed over key body landmarks, conventional univariate and multivariate data analyses are also possible. CAESAR is currently the largest and arguably the most valuable anthropometric data sources of this kind.

Anthropometric data are available via a variety of sources including technical reports, journal articles, websites and computer software packages. A list of selected anthropometric studies of different user groups is provided at the end of the chapter. Perhaps the most common software package used by E/HF professionals is the PeopleSize software by Open Ergonomics, an ergonomics consulting company based in England. PeopleSize provides an interface that allows users to pick from a large number of different user groups and anthropometric dimensions of interest (see Figure 27.6). It uses data made available through anthropometric publications and government surveys that include a large variety of body dimensions, nine nationalities, nine adult age groups and several data sets of children. Percentile values of the dimension are displayed and adjustments for clothing and postural adjustments can be made without manual calculations. The software can provide percentile estimates of an anthropometric dimension for a user group or even a combination of user groups using a weighted percentile calculation approach (see later section). It also has a Monte Carlo simulation function that allows the percentage of those accommodated and excluded to be estimated when more than one dimension is considered simultaneously.

### **DETERMINE THE PERCENTAGE OF THE TARGET USERS POPULATION TO BE ACCOMMODATED BY DESIGN**

Environments and products should be designed to accommodate the body sizes and abilities of the largest possible percentage of target users. Because there is often such a wide variety of body sizes and abilities in most user populations, it is very difficult and not economically feasible to design an environment or product that will be easily usable by everyone all of the time. Consider doorway heights. The standard doorway height in the United States is 203 cm (6'8"). While this height affords clearances that accommodate the vast majority of the public, there are still those who need to bend their neck and/or torso to avoid striking their heads on doorways. It would not be technically and economically feasible to accommodate everyone's height in the design of all doorways in today's buildings.



Conventional ergonomics and human factors practice often selects a large majority of the target user population to accommodate (e.g. 99% or 95%). But these percentages are somewhat arbitrary, perhaps due in part because anthropometric data are often reported for only a small subset of percentile values that usually include 95th and 99th percentiles. In some cases, such as those that have serious safety consequences, it may be critical to accommodate as close to 100% of the target user population as possible, while in other cases that are less critical it may be necessary to accommodate less than 95% of the population. The engineer or designer should instead weigh the benefits of accommodating larger percentages of target users (e.g. larger market, better product or environment) with the associated costs (e.g. additional design and material costs).

## UNIVARIATE ANALYSES TO INFORM DESIGN

The portion of the distribution that will be placed at a disadvantage by one design dimension will usually be the largest and/or smallest values of the distribution. For example, limitations in the heights and widths of office furniture would potentially exclude some of the tallest or widest of a user population. Limitations in weight tolerance such as in certain playground equipment might exclude the heaviest of the target user population. Grip force requirements and reach requirements might exclude those with the most limited gripping strength and reaching abilities. A height-adjustable seat might be less accommodating to those with extremely long or short legs (or more accurately, large or small popliteal heights). Adjustable furniture such as height adjustable chairs is usually designed to accommodate the mid-range of a distribution in order to allow a particular range of adjustment to benefit the greatest possible percentage of the target user population.

When the distributional characteristics of the corresponding body size or function of the user population are known and follow a Gaussian distribution, the value(s) of one design dimension (e.g. a height, width or force or the upper and lower critical values of the dimension) that would accommodate given percentage of the population can be determined with the following equation:

$$\text{Critical value of percentile}_{\text{above the mean}} = X_{\text{mean}} + (Z \text{ score}_{\text{selected critical value in distribution}} * S.D.)$$

$$\text{Critical value of percentile}_{\text{below the mean}} = X_{\text{mean}} - (Z \text{ score}_{\text{selected critical value in distribution}} * S.D.)$$

Take a hypothetical example for which the mean of an anthropometric dimension for a user population is 100 cm with a standard deviation of 10 cm. Sample critical percentile values for accommodating 95% of the upper, lower and middle portions of the distribution are as follows:

*Upper 95%:*

$$5\text{th percentile} = 100 - (1.65 * 10) = 83.5 \text{ cm} \quad 83.5 \text{ cm or greater}$$

*Lower 95%:*

$$95\text{th percentile} = 100 + (1.65 * 10) = 116.5 \text{ cm} \quad 116.5 \text{ cm or lesser}$$

*Middle 95%:*

$$2.5\text{th percentile} = 100 - (1.96 * 10) = 80.4 \text{ cm}$$

$$97.5\text{th percentile} = 100 + (1.96 * 10) = 119.6 \text{ cm} \quad \text{between } 80.4 \text{ and } 119.6 \text{ cm}$$

## MULTIVARIATE ANALYSES TO INFORM DESIGN

Multivariate analyses include the use of correlation and regression analyses, as well as human modelling methods that can provide more accurate, but more complex, analyses for evaluating

the extent that a design accommodates a population. The correlation between two dimensions provides a measure of how strongly two dimensions co-vary linearly. Ergonomic design accommodations require multiple approaches to satisfy the complex requirements of work-related tasks. Several research groups have investigated univariate and multivariate approaches to accommodate anthropometric variability in ergonomic design (e.g. Garneau and Parkinson, 2011; Human Factors and Ergonomics Society 300 Committee, 2003; Paquet and Feathers, 2004; Parkinson and Reed, 2006; Reed et al., 1999).

Researchers also have presented methods for multivariate anthropometry data analysis (e.g. Case et al., 2009; Robinette, 1998). Static and dynamic digital human modelling approaches require manipulating models of various sizes in virtual environments to assess the person-design fit. Analyses that require evaluating the fit of 2 or 3 anthropometric dimensions against design criteria can rely on a boundary case analysis, an approach in which actual cases (i.e. people within the database) are selected to represent a boundary (i.e. limit of a combination of anthropometric variables). These analyses require the raw anthropometric data for 2–3 anthropometric dimensions to analyse the data statistically or to visualise in 3-D to identify the solution space for ‘fit’ versus ‘non-fit’. Those ‘cases’ having the dimensions that are within the boundaries are ‘fit’, and the cases that fall outside of the boundary defined by the anthropometric dimensions are ‘non-fit’. There are several digital human models that are commercially evaluable to the designer to assess ‘fit’; Chapter 8 further describes the applications and benefits of human modelling approaches.

### NECESSARY ADJUSTMENTS FOR ENHANCED USABILITY

While anthropometric information can be used to inform design, it alone cannot predict the degree to which an environment or product will be used effectively and easily by the target user population.

Anthropometric measurements often require individuals to be measured in very little or tight fitting clothing while in standardised postures to help ensure reliable body measurements and reduce measurement variability that would otherwise be introduced due to clothing or postural changes with time. However, clothing characteristics and postural requirements should be accounted for when designing environments, tools and equipment. Some example adjustments for clothing are 2.5 cm for heights for males and 1.5 cm for females when shoes are worn, 0.8 cm for breadths and 1.5 cm for circumferences when individuals are clothed, 2.0 lb for shoes and 1.0 lb for clothing (Marras and Kim, 1993). Accounting for postural changes might include reducing anthropometric measurement values recorded when individuals are seated or standing erect to better estimate body positions when postures are relaxed or ‘slumped’.

### OTHER CONSIDERATIONS

It is beyond the scope of this chapter to summarise the methods used to estimate body dimensions from other dimensions when they do not exist, but it can be done. Additionally, there are statistical methods that can be used to weight user group populations to proportionally match the demographics of a user group. The Human Factors and Ergonomics Society has published (HFES, 2003) a series of useful guidelines for the appropriate application of anthropometry to product design. The document contains many examples to assist the designer or human factors and ergonomics professional in univariate and multivariate anthropometric applications to design. A case study is provided later in this chapter to provide one illustrative example of how such approaches can be used.

While anthropometric data often provide very precise information about the body sizes and functional abilities of user groups, exactly how individuals may interact with a design cannot be predicted with the same degree of precision. The final design of a workstation, tool or a piece of equipment therefore cannot rely on an anthropometric analysis alone. Analysis of human performance using mockups and task trailing is a recommended design practice. It requires recruiting individuals with the

desired range of body size and ability, and assessing human performance during the simulation such as judging whether or not individuals can reach a control or easily see a display for a particular design.

## **CASE STUDY: EVALUATING OFFICE WORKSTATION CLEARANCES FOR WHEELCHAIR USERS**

The ANSI/HFES 100-2007 standard for Human Factors Engineering of Computer Workstations is a technical standard which specifies ‘acceptable applications of human factors engineering principles and practices to the design and configuration of the human-hardware interfaces in computer workstations’ (ANSI/HFES, 2007). While the document is a valuable resource to human factors engineers and designers, the recommendations in the standard are largely limited to accommodating seated able-bodied individuals.

Office furniture is an integral part of many work environments in today’s modern workplace. The need for office furniture and workstations that accommodate individuals with diverse abilities is important from a legal standpoint. Federal laws in the United States prohibit workplace discrimination against people with disabilities and require employers of 15 or more employees to provide reasonable accommodations for those with known limitations (U.S. Department of Justice, 1990). Employers covered by these laws are required, within reasonable limits, to ensure that the work environment is accessible to and usable by an employee with a disability.

Many wheeled mobility users often use their wheelchairs in conjunction with computer workstations and other working surfaces. It is therefore desirable to provide sufficient clearance space under the working surface to allow a wheeled mobility device user to work with minimal obstruction. This case study describes one approach used to determine the recommended clearances of office workstations that would accommodate a large percentage of wheeled mobility device users (D’Souza et al., 2012) using the steps given earlier.

### **TARGET USER POPULATION**

The target user populations are male and female manual and powered wheelchair users between the ages of 18 and 70 years. The intent is to accommodate wheeled mobility users with a likely age range for employment that could potentially benefit from an accessible office workstation.

### **CRITICAL DESIGN AND ANTHROPOMETRIC DIMENSIONS**

Workstation clearances are impacted by the heights, widths and depths of the spaces underneath a workstation’s work surface. The anthropometric dimensions are heights, depths and widths of key anatomical or seat landmarks (Table 27.3). The locations of the armrests were not considered in the analysis because many wheelchairs available today in the United States for personal use have adjustable and/or removable armrests.

### **APPROPRIATE SOURCE OF DATA**

The data source was 338 wheelchair users ages 18–70 selected from among a larger sample of about 500 wheeled mobility users in the United States measured as a part of an anthropometry study that included a larger age range of wheeled mobility users (Steinfeld et al., 2010).

### **PERCENTAGE OF THE TARGET USER POPULATION TO BE ACCOMMODATED BY THE DESIGN**

An analysis was performed to assess the clearance dimensions needed to accommodate 90% and 95% of the sample. This is admittedly an arbitrary percentage (as discussed previously) as an attempt to investigate what clearance dimensions would be necessary to accommodate the vast majority of wheelchair users.

**TABLE 27.3**  
**Relevant Anthropometric Dimensions and Design Considerations for Office Workstations Clearances**

| Dimension                                | Description                                                                                                | Design Relevance                                                                                      |
|------------------------------------------|------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| Knee clearance height (KCH) <sup>a</sup> | Vertical distance from the floor to the higher of the superior aspect of either the knee or thigh          | Height of the clearance space at the edge of the work surface closest to the operator                 |
| Knee clearance depth (KCD) <sup>a</sup>  | Horizontal distance from the anterior aspect of the abdomen to the distal aspect of the knee               | Depth of the clearance space at knee height from the edge of the work surface closest to the operator |
| Foot clearance height (FCH) <sup>a</sup> | Vertical distance measured from the floor plane to the dorsal aspect of the foot                           | Height of the clearance space at the level of the foot                                                |
| Foot clearance depth (FCD) <sup>a</sup>  | Horizontal distance from the anterior aspect of the abdomen to the dorsal crease of the lower leg and foot | Depth of the clearance space at the level of the foot from the closest edge of the work surface       |
| Abdomen depth (AD) <sup>a</sup>          | Horizontal distance from the anterior aspect of the abdomen to the distal-most aspect of the foot          | Overall depth of the clearance space at floor level from the closest edge of the work surface         |
| Occupied width (OW)                      | Greater of the horizontal distance between the right- and leftmost aspect of either the seat or the knees  | Width of the clearance space                                                                          |

<sup>a</sup> Larger of the dimension value measured on the right and left side of the body.

### PORTION OF THE DISTRIBUTION THAT WILL BE EXCLUDED AND ACCOMMODATED

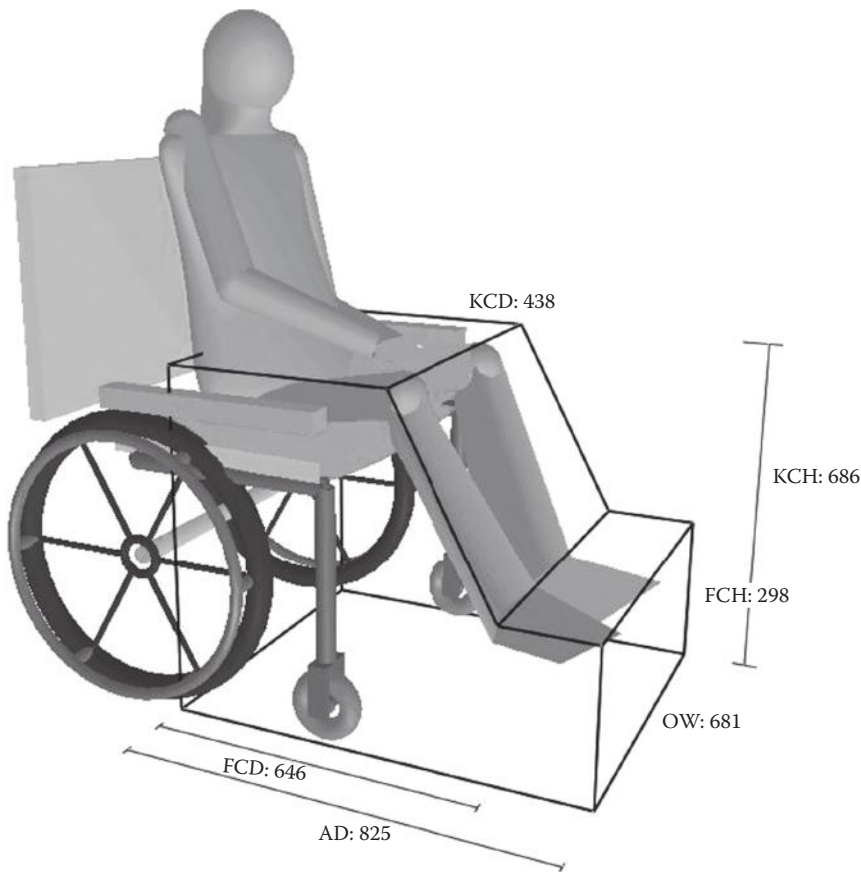
Those with largest knee and foot clearances, largest abdomen depths and largest widths of the user will not be accommodated when the size of the clearances is inadequate. Smaller individuals can fit in a large clearance space. Therefore, values for 90th and 95th percentile values of clearance were used to establish the design specifications that would accommodate 90% and 95% of the user populations, respectively.

### ANALYSIS

The 6-D values were computed for every individual in the sample, and stratified into four user groups: male and female users of manual wheelchairs and male and female users of powered wheelchairs. Two approaches were used to compute the clearance envelopes for each of the four groups: univariate and multivariate. The univariate approach estimated the percentage of the population for which the user group fits within each dimension individually. The multivariate approach estimated the percentage of the user group who would fit within the spaces when the dimensions are considered simultaneously.

*Univariate percentile approach:* For each dimension, the 90th and 95th percentile values were computed for each of the subgroups. The 90th and 95th percentile clearance envelopes were constructed directly from the corresponding percentile values of each of the dimensions for each subgroup.

*Iterative multivariate approach:* A fairly simple iterative method was used to obtain an estimate of the minimum clearance space required to accommodate a specific proportion of the sample on all of the six dimensions similar to a virtual fitting trial. A 3-D static model of each person included in the analysis was developed (Figure 27.7). Starting from the maximum observed dimension values as the reference envelope, each of the six dimension values in the reference envelope were successively decreased by one unit of the same scale, with the percentage of the sample accommodate



**FIGURE 27.7** Example clearance envelopes (mm) developed from 3-D information of wheeled mobility users.

being computed at each decrement. The percentage of the sample accommodated was computed by checking the dimension values for every subject against the reference values. A subject was considered 'accommodated' if all of the measurements were less than the corresponding reference value. The process was repeated until the minimum dimensions of the bounding envelope required to accommodate a set proportion of the sample (e.g. 90% or 95%) were identified.

Tables 27.4 and 27.5 summarise the results of the univariate and multivariate analyses. The results demonstrated that the clearance requirements were much larger for men than for women. The analysis also demonstrated that the multivariate iterative approach to assessing clearances results in larger space requirements. This is because body dimensions are not correlated perfectly (e.g. it is possible to have a person with low knee height but large width). Therefore, the multivariate approach provides a better estimate of the percentage of individuals who will be accommodated or excluded by design.

The results also demonstrate that the ANSI/HFES 100-2007 prescribed clearance space for reclined seating are not large enough to accommodate the desired percentage of the male wheelchair user population. For example, the 95th percentile clearance space for male manual wheelchair users exceeded the recommended guidelines in each of the six design dimensions, with an extremely large mismatch between the guidelines and corresponding values of foot clearance height (Figure 27.8). Much larger clearance envelopes would therefore be required to accommodate the clearance requirements of 95% of male manual wheelchair users. In addition to the clearance space for wheelchair users being wider and higher, the comparison shows the need for substantially greater depth and height, particularly at the level of the foot.

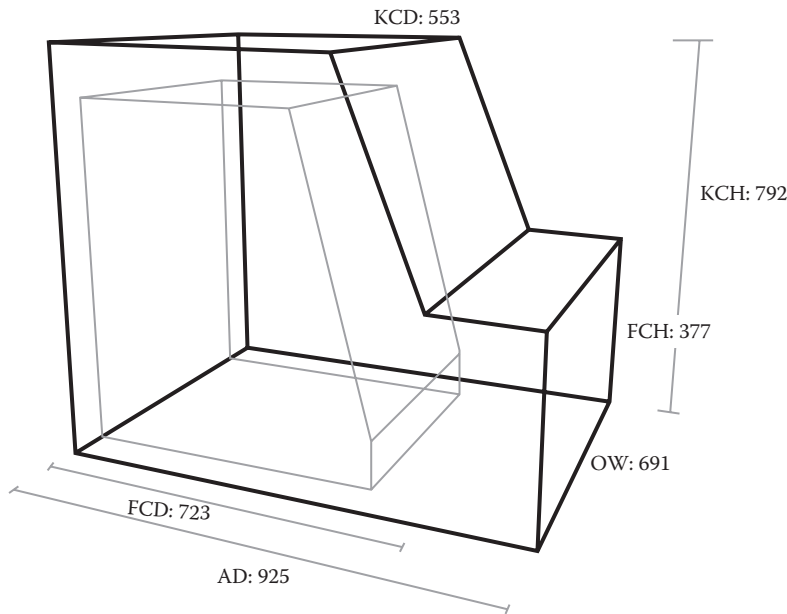
TABLE 27.4  
Clearance Envelope Dimensions in mm for Manual Wheelchair Users

| TABLE 27.4<br>Clearance Envelope Dimensions in mm for Manual Wheelchair Users |                                        |            |        |                    |        |                     |                                         |            |        |                    |        |                     |
|-------------------------------------------------------------------------------|----------------------------------------|------------|--------|--------------------|--------|---------------------|-----------------------------------------|------------|--------|--------------------|--------|---------------------|
| Dimension                                                                     | Male Manual Wheelchair Users (n = 105) |            |        |                    |        |                     | Female Manual Wheelchair Users (n = 81) |            |        |                    |        |                     |
|                                                                               | Observed<br>Median                     | Univariate |        | Iterative Solution |        | Observed<br>Maximum | Observed<br>Median                      | Univariate |        | Iterative Solution |        | Observed<br>Maximum |
|                                                                               |                                        | 90%ile     | 95%ile | 90%ile             | 95%ile |                     |                                         | 90%ile     | 95%ile | 90%ile             | 95%ile |                     |
|                                                                               |                                        |            |        |                    |        |                     |                                         |            |        |                    |        |                     |
| KCH                                                                           | 666                                    | 713        | 737    | 761                | 792    | 793                 | 617                                     | 680        | 698    | 709                | 714    | 715                 |
| KCD                                                                           | 384                                    | 466        | 510    | 522                | 553    | 554                 | 321                                     | 423        | 448    | 498                | 503    | 504                 |
| FCH                                                                           | 209                                    | 303        | 321    | 346                | 377    | 378                 | 222                                     | 310        | 329    | 433                | 438    | 439                 |
| FCD                                                                           | 481                                    | 586        | 676    | 693                | 723    | 724                 | 424                                     | 570        | 634    | 662                | 667    | 668                 |
| AD                                                                            | 627                                    | 745        | 816    | 895                | 925    | 926                 | 575                                     | 724        | 750    | 825                | 830    | 830                 |
| OW                                                                            | 420                                    | 643        | 665    | 661                | 691    | 692                 | 445                                     | 669        | 705    | 742                | 747    | 747                 |



TABLE 27.5  
Clearance Envelope Dimensions in mm for Powered Wheelchair Users

| TABLE 27.5<br>Clearance Envelope Dimensions in mm for Powered Wheelchair Users |                                        |            |        |                    |        |                     |                                          |            |        |                    |     |                     |
|--------------------------------------------------------------------------------|----------------------------------------|------------|--------|--------------------|--------|---------------------|------------------------------------------|------------|--------|--------------------|-----|---------------------|
| Dimension                                                                      | Male Powered Wheelchair Users (n = 81) |            |        |                    |        |                     | Female Powered Wheelchair Users (n = 65) |            |        |                    |     |                     |
|                                                                                | Observed<br>Median                     | Univariate |        | Iterative Solution |        | Observed<br>Maximum | Observed<br>Median                       | Univariate |        | Iterative Solution |     | Observed<br>Maximum |
|                                                                                |                                        | 90%ile     | 95%ile | 90%ile             | 95%ile |                     |                                          | 90%ile     | 95%ile |                    |     |                     |
|                                                                                |                                        |            |        |                    |        |                     |                                          |            |        |                    |     |                     |
| KCH                                                                            | 691                                    | 747        | 759    | 768                | 795    | 796                 | 666                                      | 718        | 726    | 727                | 746 | 747                 |
| KCD                                                                            | 367                                    | 508        | 532    | 619                | 646    | 647                 | 315                                      | 394        | 440    | 454                | 473 | 474                 |
| FCH                                                                            | 261                                    | 359        | 384    | 458                | 485    | 486                 | 279                                      | 380        | 408    | 434                | 452 | 453                 |
| FCD                                                                            | 479                                    | 619        | 719    | 722                | 749    | 750                 | 423                                      | 539        | 590    | 679                | 697 | 698                 |
| AD                                                                             | 619                                    | 786        | 836    | 830                | 857    | 858                 | 556                                      | 695        | 732    | 872                | 891 | 891                 |
| OW                                                                             | 464                                    | 613        | 631    | 667                | 694    | 695                 | 449                                      | 626        | 678    | 719                | 738 | 738                 |



**FIGURE 27.8** Comparison of the ANSI/HFES 100-2007 prescribed clearance space for reclined seating with thin grey lines and the 95th percentile clearances for space with thick black lines for male manual wheelchair users. The graphical representation shows dimensions for ANSI/HFES 100-2007 reclined seating with grey lines and the 95th percentile clearances for space with thick black lines.

### NECESSARY ADJUSTMENTS TO ENHANCE USABILITY

The anthropometric database used for the analysis allowed individuals to be fully clothed and seated in a ‘relaxed’ posture that was assumed to approximate a comfortable working posture. Therefore, no adjustments were made to the data for this analysis. The analysis provided only an initial assessment of the appropriate clearances needed for workstation furniture to accommodate the body and chair sizes of wheelchair users. Mock-ups and user trialling of wheelchair users would be recommended to ensure that the clearance dimensions allow wheelchair users to efficiently and effectively access and use the workstations.

### CONCLUSIONS

This chapter has provided the reader with an introduction to anthropometric methods and some guidelines about how to apply anthropometry to inform the ergonomic design of products, equipment and environments. New data sets, approaches to the collection of anthropometric data, multivariate statistical analyses and human modelling methods now provide opportunities to apply anthropometry to design in a more accurate and meaningful way. Anthropometry provides a ‘starting point’ to help ensure designs are ergonomically sound for the intended users, but does not replace other types of approaches such as usability testing to ensure that products, equipment and environments can be used effectively by the intended user groups.

### REFERENCES

- Aldridge, K., Boyadjiev, S.A., Capone, G.T., DeLeon, V.B., and Richtsmeier, J.T. (2005). Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *American Journal of Medical Genetics A*, **138**: 247–253.
- ANSI/HFES (2007). ANSI/HFES 100–2007 Human factors engineering of computer workstations.

- Anthropology Research Staff, Ed. (1978a). *Anthropometric Source Book Volume 1: Anthropometry for Designers* (NASA Reference Publication 1024). Houston, TX: NASA Scientific and Technical Information Office.
- Anthropology Research Staff, Ed. (1978b). *Anthropometric Source Book Volume 2: A Handbook of Anthropometric Data* (NASA Reference Publication 1024). Houston, TX: NASA Scientific and Technical Information Office.
- Bear-Lehman, J., Miller, P.A., Adler, M., Buonocore, J.M., Coles, N., Kneafsey, B.S., Katz-Sillman, F., and Sherman-Amsel, H. (2003). An exploration of hand strength and sensation in community elders. *Topics in Geriatric Rehabilitation*, **19**(2): 127–136.
- Boatright, J.R., Kiebzak, G.M., O'Neil, D.M., and Peindl, R.D. (1997). Measurement of thumb abduction strength: Normative data and a comparison with grip and pinch strength. *The Journal of Hand Surgery*, **22**(5): 843–848.
- Bohannon, R.W., Peolsson, A., Massy-Westropp, N., Desrosiers, J., and Bear-Lehman, J. (2006). Reference values for adult grip strength measured with a Jamar dynamometer: a descriptive meta-analysis. *Physiotherapy*, **92**(1): 11–15.
- Brennan, P., Bohannon, R.W., Pescatello, L.S., Marschke, L., Hasson, S., and Murphy, M. (2004). Grip strength norms for elderly women. *Perceptual and Motor Skills*, **99**: 899–902.
- Case, K., Marshall, R., Hogberg, D., Summerskill, S., Gyi, D., and Sims, R. (2009). HADRIAN: Fitting trials by digital human modelling. In: *Digital Human Modeling*, Berlin, Germany, Springer, pp. 673–680.
- Chaffin, D.B. (2001). *Digital Human Modeling for Vehicle and Workspace Design*. Warrendale, PA: SAE, Inc.
- Chaffin, D., Andersson, G., and Martin, B. (2006). *Occupational Biomechanics*, 4th edn. New York, Wiley.
- Brooke-Wavell, K., Jones, P.R.M., and West, G.M. (1994). Reliability and repeatability of 3-D body scanner (LASS) measurements compared to anthropometry. *Annals of Human Biology*, **21**: 571–577.
- Clauser, C.T.I., Bradtmiller, B., McConville, J., and Gordon, C. (1988). Measurer's Handbook: U.S. Army Anthropometric Survey 1987–1988. Technical Report NATICK/TR-88/043. United States Army NATICK Research, Development and Engineering Center, Natick, MA.
- Crosby, C.A., Wehbe, M.A., and Mawr, B. (1994). Hand strength: Normative values. *Journal of Hand Surgery*, **19A**: 665–670.
- Damon, A., Stoudt, H.W., and McFarland, R.A. (1966). *The Human Body in Equipment Design*. Cambridge, MA: Harvard University Press.
- D'Souza, C., Paquet, V., and Steinfeld, E. (2012). Clearance envelopes of wheeled mobility device users for computer workstations. In: *Proceedings of the Human Factors and Ergonomics Society*, Boston, MA, October 22–26.
- Desrosiers, J., Bravo, G., Hebert, R., and Mercier, L. (1995a). Normative data for grip strength of elderly men and women. *American Journal of Occupational Therapy*, **49**: 637–644.
- Feathers, D., Paquet, V., and Drury, C. (2004). Measurement consistency and three-dimensional electromechanical anthropometry. *International Journal of Industrial Ergonomics*, **33**(3): 181–190.
- Fess, E.E. (1992). Grip strength. In: *Clinical Assessment Recommendations*, J.S. Casanova (ed.), Chicago, IL, American Society of Hand Therapists, pp. 41–45.
- Fiebert, I.M., Roach, K.E., Armstrong, T., Mandel, D.W., and Donohue, M. (1995). Dynamometric grip strength assessment of subjects sixty years and older. *Physical & Occupational Therapy in Geriatrics*, **13**(4): 27–40.
- Floyd, W.F., Guttmann, L., Noble, C.W., Parkes, K.R., and Ward, J. (1966). A study of the space requirements of wheelchair users. *Paraplegia*, **4**: 24–37.
- Garneau, C.J. and Parkinson, M.B. (2011). A comparison of methodologies for designing for human variability. *Journal of Engineering Design*, **22**(7): 505–521.
- Hanten, W.P., Chen, W.-Y., Austin, A.A., Brooks, R.E., Carter, H.C., and Law, C.A. (1999). Maximum grip strength in normal subjects from 20 to 64 years of age. *Journal of Hand Therapy*, **12**: 193–200.
- HFES (Human Factors and Ergonomics Society) 300 Committee. (2003). *Guidelines for Using Anthropometric Data in Product Design*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Horowitz, B.P., Tollin, R., and Cassidy, G. (1997). Grip strength: Collection of normative data with community dwelling elderly. *Physical and Occupational Therapy in Geriatrics*, **15**(1): 53–64.
- Human Factors and Ergonomics Society 300 Committee. (2003). *Guidelines for Using Anthropometric Data in Product Design*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Imrhan, S.N. and Loo, C.H. (1989a). Modelling wrist-twisting strength of the elderly. *Ergonomics*, **31**(2): 1807–1819.
- Imrhan, S.N. and Loo, C.H. (1989b). Trends in finger pinch strength in children, adults, and the elderly. *Human Factors*, **31**(6): 689–701.

- International Standardisation Organisation, Ed (1996). *Basic Human Body Measurements for Technical Design* (ISO Standard 7250). Geneva, Switzerland: International Organisation for Standardisation.
- International Standardisation Organisation, Ed (2003). *General Requirements for Establishing an Anthropometric Database* (ISO Standard 15535). Geneva, Switzerland: International Organisation for Standardisation.
- Kennedy, K. (1986). A collation of United States Air Force Anthropometry (U) (AAMRL-TR-85-062). Aerospace Medical Research Laboratory, Aiero Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, OH.
- Kouchi, M. and Mochimaru, M. (2011). Errors in landmarking and the evaluation of the accuracy of traditional and 3D anthropometry. *Applied Ergonomics*, **42**: 518–527.
- Kulkarni, D.S.R., Chitodkar, V., Gurjar, V., Ghaisas, C.V., and Mannikar, A.V. (2011). SIZE INDIA- Anthropometric size measurement of Indian driving population. SAE Technical Paper 2011-26-0108. SAE International, Warrendale, PA.
- Lohman, T., Roche, A., and Martorell, R. (1988). *Anthropometric Standardisation Reference Manual*. Champaign, IL: Human Kinetics Publishers.
- Marras, W. and Kim, J. (1993). Anthropometry of industrial populations. *Ergonomics*, **36**(4): 371–377.
- Marshall, R., Case, K., Gyi, D., Oliver, R., and Porter, J.M. (2002). HADRIAN: An integrated design ergonomics analysis tool. In: *Proceedings of the XVI Annual International Occupational Ergonomics and Safety Conference*, Toronto, Ontario, Canada.
- Marshall, R., Case, K., Porter, J.M., Summerskill, S., Gyi, D.E., Davis, P., and Sims, R. (2010). HADRIAN: A virtual approach to design for all. *Journal of Engineering Design*, **21**(2–3): 253–273.
- Mathiowetz, V., Kashman, N., Volland, G., Weber, K., Dowe, M., and Rogers, S. (1985b). Grip and pinch strength: Normative data for adults. *Archives of Physical Medicine and Rehabilitation*, **66**: 69–72.
- Morant, G.M. (1947). Anthropometric problems in the Royal Air Force. *British Medical Bulletin*, **5**(1017): 25–31.
- Paquet, V. and Feathers, D. (2004). An anthropometric study of manual and powered wheelchair users. *International Journal of Industrial Ergonomics*, **33**(3): 191–204.
- Parkinson, M. and Reed, M.P. (2006). Optimizing vehicle occupant packaging. *SAE Transactions: Journal of Passenger Cars—Mechanical Systems*, **115**(6): 890–901.
- Peebles, L. and Norris, B. (1998). *Adultdata: The Handbook of Adult Anthropometrics and Strength Measurements—Data for Design Safety*. London, U.K.: Department of Trade and Industry.
- Pheasant, S. and Haslegrave, C.M. (2006). *Bodyspace: Anthropometry, Ergonomics, and the Design of Work*. Boca Raton, FL: CRC Press, Taylor and Francis Group.
- Reed, M., Manary, M., and Schneider, L. (1999). Methods for measuring and representing automobile occupant posture. SAE Technical Paper 1999-01-0959.
- Reed, M.P., Parkinson, M.B., and Klinkenberger, A.L. (2003). Assessing the validity of kinematically generated reach envelopes for simulations of vehicle operators. SAE Technical Paper 2003-01-2216. *SAE Transactions: Journal of Passenger Cars—Mechanical Systems*, **112**. doi:10.4271/2003-01-2216.
- Research Institute of Human Engineering for Quality Life. (1997). Japanese body size data 1992–1994. Osaka, Japan [in Japanese].
- Rice, C.L., Cunningham, D.A., Paterson, D.H., and Rechnitzer, P.A. (1989). Strength in an elderly population. *Archives of Physical Medicine and Rehabilitation*, **70**: 391–397.
- Robinette, K.M. (1998). Multivariate methods in engineering anthropometry. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, Chicago, IL, October 5–9. The Human Factors and Ergonomics Society, Santa Monica, CA, pp. 719–721.
- Robinette, K.M. (2000). CAESAR measures up. *Ergonomics in Design*, **8**(3): 17–23.
- Shechtman, O., Mann, W.C., Justiss, M.D., and Tomita, M. (2004). Grip strength in the frail elderly. *American Journal of Physical Medicine and Rehabilitation*, **83**(11): 819–826.
- Steinfeld, E. (1986). *Hands-on Architecture: Volume 3, Parts 1 & 2*. Buffalo, NY: Center for Inclusive Design and Environmental Analysis, School of Architecture and Planning, University at Buffalo.
- Steinfeld, E., Paquet, V., D'Souza, C., Joseph, C., and Maisel, J. (2010). Anthropometry of Wheeled Mobility Project. Submitted to U.S. Access Board, Washington, DC [Final Report]. [http://www.udeworld.com/documents/anthropometry/pdfs/AnthropometryofWheeledMobilityProject\\_FinalReport.pdf](http://www.udeworld.com/documents/anthropometry/pdfs/AnthropometryofWheeledMobilityProject_FinalReport.pdf).
- U.S. Department of Justice. (1990). Americans with Disabilities Act of 1990, As Amended 2008. U.S. Department of Justice, Washington, DC. Retrieved from <http://www.ada.gov/pubs/ada.htm>.
- Voorbij, A.I.M. and Steenbekkers, L.P.A. (2002). The twisting force of aged consumers when opening a jar. *Applied Ergonomics*, **33**(1): 105–109.

## Selected Anthropometric Data Sources and Reference Materials

### *Anthropometry: Aging*

- Annis, J.F., Case, H.W., Clauser, C.E., and Bradtmiller, B. (1991). Anthropometry of an aging work force. *Experimental Aging Research*, **17**: 157–176.
- Annis, J.F. (1996). Aging effects on anthropometric dimensions important to workplace design. *International Journal of Industrial Ergonomics*, **18**(5–6): 381–388.
- Bassey, E.J. and Harries, U.J. (1993). Normal values for hand grip strength in 920 men and women over 65 years, and longitudinal changes over 4 years in 620 survivors. *Clinical Science*, **84**: 331–337.
- Bear-Lehman, J., Miller, P.A., Adler, M., Buonocore, J.M., Coles, N., Kneafsey, B.S., Katz-Sillman, F., and Sherman-Amsel, H. (2003). An exploration of hand strength and sensation in community elders. *Topics in Geriatric Rehabilitation*, **19**(2): 127–136.
- Brennan, P., Bohannon, R.W., Pescatello, L.S., Marschke, L., Hasson, S., and Murphy, M. (2004). Grip strength norms for elderly women. *Perceptual and Motor Skills*, **99**: 899–902.
- Damon, A. and Stout, H.W. (1963). The functional anthropometry of old men. *Human Factors*, **5**: 485–491.
- Desrosiers, J., Bravo, G., Hebert, R., and Mercier, L. (1995a). Normative data for grip strength of elderly men and women. *American Journal of Occupational Therapy*, **49**: 637–644.
- Desrosiers, J., Bravo, G., Hebert, R., and Mercier, L. (1995b). Impact of elbow position on grip strength of elderly men. *Journal of Hand Therapy*, **8**: 27–30.
- Fiebert, I.M., Roach, K.E., Armstrong, T., Mandel, D.W., and Donohue, M. (1995). Dynamometric grip strength assessment of subjects sixty years and older. *Physical & Occupational Therapy in Geriatrics*, **13**(4): 27–40.
- Horowitz, B.P., Tollin, R., and Cassidy, G. (1997). Grip strength: Collection of normative data with community dwelling elderly. *Physical and Occupational Therapy in Geriatrics*, **15**(1): 53–64.
- Imrhan, S.N. and Loo, C.H. (1989a). Modelling wrist-twisting strength of the elderly. *Ergonomics*, **31**(2): 1807–1819.
- Molenbroek, J.F.M. (1987). Anthropometry of elderly people in the Netherlands; research and applications. *Applied Ergonomics*, **18**(3), 187–199.
- Rice, C.L., Cunningham, D.A., Paterson, D.H., and Rechnitzer, P.A. (1989). Strength in an elderly population. *Archives of Physical Medicine and Rehabilitation*, **70**: 391–397.
- Shechtman, O., Mann, W.C., Justiss, M.D., and Tomita, M. (2004). Grip strength in the frail elderly. *American Journal of Physical Medicine and Rehabilitation*, **83**(11): 819–826.
- Voorbij, A.I.M. and Steenbekkers, L.P.A. (2002). The twisting force of aged consumers when opening a jar. *Applied Ergonomics*, **33**(1): 105–109.

### *Anthropometry: Children*

- Donoso, H. (1987). Anthropometry, body composition and maximum aerobic strength in a chosen sample of Chilean school children aged 7–17: A transversal study. *Archivos de la Sociedad Chilena de Medicina del Deporte*, **32**: 42–48.
- Feathers, D., Pavlovic-Veselinovic, S., and Hedge, A. (2013). Measures of fit and discomfort for elementary school children in Serbia. *Work*, **44**: S73–S81.
- Snyder, R.G., Schneider, L.W., Owings, C.L., Reynolds, H.M., Golomb, D.H., and Schork, M.A. (1977). Anthropometry of infants, children, and youths to age 18 for product safety design. Highway Safety Research Institute, The University of Michigan, Ann Arbor, Michigan [Final Report].
- Steenbekkers, L. and Molenbroek, J. (1990). Anthropometric data of children for non-specialized users. *Ergonomics*, **33**(4): 421–429.

### *Anthropometry: Data Treatment, Methods, and Analysis*

- Ahlstrom, V. and Longo, K. (2003). *Human Factors Design Standard* (HF-STD-001). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.
- Aldridge, K., Boyadjiev, S.A., Capone, G.T., DeLeon, V.B., and Richtsmeier, J.T. (2005). Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *American Journal of Medical Genetics A*, **138**: 247–253.
- Feathers, D., Paquet, V., and Drury, C. (2004). Measurement consistency and three-dimensional electromechanical anthropometry. *International Journal of Industrial Ergonomics*, **33**(3): 181–190.
- Garneau, C.J. and Parkinson, M.B. (2011). A comparison of methodologies for designing for human variability. *Journal of Engineering Design*, **22**(7): 505–521.
- ISO 20685:2005. (2005). 3-D scanning methodologies for internationally compatible anthropometric databases. International Standards Organization (ISO), Geneva, Switzerland.

- ISO 7250-1:2007. (2007). Basic human body measurements for technological design. International Standards Organization (ISO), Geneva, Switzerland.
- Kouchi, M. and Mochimaru, M. (2011). Errors in landmarking and the evaluation of the accuracy of traditional and 3D anthropometry. *Applied Ergonomics*, **42**: 518–527.
- Morant, G.M. (1947). Anthropometric problems in the Royal Air Force. *British Medical Bulletin*, **5** (1017): 25–31.
- Paquet, V. and Feathers, D. (2004). An anthropometric study of manual and powered wheelchair users. *International Journal of Industrial Ergonomics*, **33**(3): 191–204.
- Parkinson, M. and Reed, M.P. (2006). Optimizing vehicle occupant packaging. *SAE Transactions: Journal of Passenger Cars—Mechanical Systems*, **115**(6): 890–901.
- Pheasant, S. and Haslegrave, C.M. (2006). *Bodyspace: Anthropometry, Ergonomics and the Design of Work*, 3rd edn. Boca Raton, FL: CRC Press, Taylor and Francis Group.
- Reed, M.P., Parkinson, M.B., and Klinkenberger, A.L. (2003). Assessing the validity of kinematically generated reach envelopes for simulations of vehicle operators. SAE Technical Paper 2003-01-2216. *SAE Transactions: Journal of Passenger Cars—Mechanical Systems*, **112**.
- Roebuck, J.A., Kroemer, K.H.E., and Thomson, W.G. (1975). *Engineering Anthropometry Methods*. New York: Wiley.
- Richards, L. and Palmiter-Thomas, P. (1996). Grip strength measurement: A critical review of tools, methods, and clinical utility. *Critical Reviews in Physical and Rehabilitation Medicine*, **8**(1&2): 87–109.
- VDI/VDE 2617 part 6.2:2005. (2005). Accuracy of coordinate measuring machines characteristics and their testing guideline for the application of DIN EN ISO 10360 to coordinate measuring machines with optical distance sensors.
- Anthropometry: Large Data Sets*
- Anthropology-Research-Project, Ed (1978). *Anthropometric Source Book* (NASA Reference Publication 1024) (Vols. I, II, III). Washington, DC: NASA, Science and Technical Information Office.
- Gordon, C., Bradtmiller, B., Clauser, C., Churchill, T., McConville, J., Tebbetts, I., and Walker, R. (1989). 1987–1988 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics. Technical Report (TR-89/027) (AD A209600). U.S. Army Natick Research, Development and Engineering Center, Natick, MA.
- Kennedy, K. (1986). A collation of United States Air Force Anthropometry (U) (AAMRL-TR-85-062). Aerospace Medical Research Laboratory, Aiero Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, OH.
- Kulkarni, D.S.R., Chitodkar, V., Gurjar, V., Ghaisas, C.V., and Mannikar, A.V. (2011). SIZE INDIA—Anthropometric size measurement of Indian driving population. SAE Technical Paper 2011-26-0108. SAE International, Warrendale, PA.
- Marras, W. and Kim, J. (1993). Anthropometry of industrial populations. *Ergonomics*, **36**(4): 371–377.
- Pett, L.B. and Ogilvie, G.F. (1957). The Canadian weight-height survey. *Human Biology*, **28**: 177–188.
- Research Institute of Human Engineering for Quality Life (1997). Japanese Body Size Data 1992–1994. Osaka, Japan [in Japanese].
- Robinette, K.M. (2000). CAESAR measures up. *Ergonomics in Design*, **8**(3): 17–23.
- Anthropometry: Static and Dynamic Strength*
- Balogun, J.A., Akomolafe, C.T., and Amusa, L.O. (1991). Grip strength: Effects of testing posture and elbow position. *Archives of Physical Medicine and Rehabilitation*, **72**: 280–283.
- Boatright, J.R., Kiezbak, G.M., O’Neil, D.M., and Peindl, R.D. (1997). Measurement of thumb abduction strength: Normative data and a comparison with grip and pinch strength. *The Journal of Hand Surgery*, **22**(5): 843–848.
- Bohannon, R.W., Peolsson, A., Massy-Westropp, N., Desrosiers, J., and Bear-Lehman, J. (2006). Reference values for adult grip strength measured with a Jamar dynamometer: A descriptive meta-analysis. *Physiotherapy*, **92**(1): 11–15.
- Crosby, C.A., Wehbé, M.A., and Mawr, B. (1994). Hand strength: Normative values. *Journal of Hand Surgery*, **19A**: 665–670.
- Fess, E.E. (1992). Grip strength. In: *Clinical Assessment Recommendations*, J.S. Casanova (ed.). Chicago, IL: American Society of Hand Therapists, pp. 41–45.
- Hanten, W.P., Chen, W.-Y., Austin, A.A., Brooks, R.E., Carter, H.C., and Law, C.A. (1999). Maximum grip strength in normal subjects from 20 to 64 years of age. *Journal of Hand Therapy*, **12**: 193–200.
- Kuzala, E.A. and Vargo, M.C. (1992). The relationship between elbow position and grip strength. *American Journal of Occupational Therapy*, **46**: 509–512.



- Mathiowetz, V., Kashman, N., Volland, G., Weber, K., Dowe, M., and Rogers, S. (1985b). Grip and pinch strength: Normative data for adults. *Archives of Physical Medicine and Rehabilitation*, **66**: 69–72.
- Mathiowetz, V., Rennells, C., and Donahoe, L. (1985a). Effect of elbow position on grip and key pinch strength. *The Journal of Hand Surgery*, **10A**(5): 694–696.
- Mathiowetz, V., Weber, K., Volland, G., and Kashman, N. (1984). Reliability and validity of grip and pinch strength evaluations. *Journal of Hand Surgery*, **9A**(2): 222–226.
- Peebles, L. and Norris, B. (1998). *Adultdata: The Handbook of Adult Anthropometrics and Strength Measurements—Data for Design Safety*. London, U.K.: Department of Trade and Industry.
- Anthropometry: Understudied Populations/Special Populations*
- Chen, C.C., Kasven, N., Karpatkin, H.I., and Sylvester, A. (2007). Hand strength and perceived manual ability among patients with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation*, **88**: 794–797.
- Chen, C.C. and Giustino, J. (2007). Grip strength, perceived ability, and health status in individuals with arthritis: An exploratory study. *Occupational Therapy in Health Care*, **21**(4): 1–18.
- D’Souza, C., Paquet, V., and Steinfeld, E. (2012). Clearance envelopes of wheeled mobility device users for computer workstations. *Proceedings of the Human Factors and Ergonomics Society*, Boston, MA, October 22–26.
- Ferraz, M., Ciconelli, R., Araujo, P., Oliveira, L., and Atra, E. (1992). The effect of elbow flexion and time of assessment on the measurement of grip strength in rheumatoid arthritis. *Journal of Hand Surgery (American)*, **17**: 1099–1103.
- Floyd, W.F., Guttmann, L., Noble, C.W., Parkes, K.R., and Ward, J. (1966). A study of the space requirements of wheelchair users. *Paraplegia*, **4**: 24–37.
- Fraser, A., Vallow, J., Preston, A., and Cooper, R.G. (1999). Predicting ‘normal’ grip strength for rheumatoid arthritis patients. *Rheumatology*, **38**: 521–528.
- Hobson, D. and Molenbroek, J. (1990). Anthropometry and design for the disabled: Experiences with seating design for cerebral palsy population. *Applied Ergonomics*, **21**(1): 43–54.
- Imrhan, S.N. and Loo, C.H. (1989b). Trends in finger pinch strength in children, adults, and the elderly. *Human Factors*, **31**(6): 689–701.
- Marshall, R., Case, K., Gyi, D., Oliver, R., and Porter, J.M. (2002). HADRIAN: An integrated design ergonomics analysis tool. In: *Proceedings of the XVI Annual International Occupational Ergonomics and Safety Conference*, Toronto, Ontario, Canada.
- Marshall, R., Case, K., Porter, J.M., Summerskill, S., Gyi, D.E., Davis, P., and Sims, R. (2010). HADRIAN: A virtual approach to design for all. *Journal of Engineering Design*, **21**(2–3): 253–273.
- Paquet, V. and Feathers, D. (2004). An anthropometric study of manual and powered wheelchair users. *International Journal of Industrial Ergonomics*, **33**: 191–204.
- Steinfeld, E. (1986). *Hands-on Architecture: Volume 3, Parts 1 & 2*. Buffalo, NY: Center for Inclusive Design and Environmental Analysis, School of Architecture and Planning, University at Buffalo.
- Steinfeld, E., Maisel, J., Feathers, D., and D’Souza, C. (2010). Anthropometry and standards for wheeled mobility: An international comparison. *Assistive Technology*, **22**(1): 51–67.
- Steinfeld, E., Paquet, V., D’Souza, C., Joseph, C., and Maisel, J. (2010). Anthropometry of Wheeled Mobility Project. Submitted to U.S. Access Board, Washington, DC [Final Report]. [http://www.udeworld.com/documents/anthropometry/pdfs/AnthropometryofWheeledMobilityProject\\_FinalReport.pdf](http://www.udeworld.com/documents/anthropometry/pdfs/AnthropometryofWheeledMobilityProject_FinalReport.pdf).



# *Section V*

---

## *Methods in Organisational and Social Human Factors*

The systems perspective now underpins much ergonomics/human factors (E/HF) design and analysis. This continues the trend that has been known as ‘macroergonomics’ (particularly in North America) and perhaps helps us as a discipline to highlight something that we have always naturally done – understand how the social and organisational context of work affects the interactions and tasks of individuals and teams.

Waterson (Chapter 28) provides an overview of sociotechnical systems theory and presents methods that can be used to analyse, plan and evaluate sociotechnical systems. He illustrates the STS approach with case studies from manufacturing and healthcare. This chapter highlights, as in other parts of the book, the challenges of dealing with increasingly complex and distributed workplaces, and also notes that there is a particular need to understand how the STS approach, which has been primarily driven by academic research, is more clearly adopted by practitioners. One of the strengths of E/HF is the relatively close relationship between academia and practice, with many academics (in particular, the editors of this volume!) spending most of their time on projects that directly work with industry partners; it is tremendously important that we continue to find methods and approaches that bridge between research and practice.

One of the elements that is often found in sociotechnical systems is a complex structure of teams (Marshall, Bisbey and Salas [Chapter 29]). After considering what they describe as the nine ‘C’s of teamwork, they describe the different approaches that can be taken to measure teamwork, focusing on attitudes, behaviours and cognitions. They highlight that there is increasing potential in modelling teams and their interactions, and that any methods to capture team performance should remember that both the overall team parameters and those associated with individuals within those teams are of importance.

Chapter 30 by Kirwan and Chapter 31 by Ryan cover the prediction and analysis of unusual events, such as incidents and accidents. Kirwan provides a comprehensive introduction to fundamental concepts, such as error classification, that underpin our assessment of the likelihood of incidents, and thus help us to mitigate against them. He presents specific examples of human error quantification and identification, and discusses the relationship between human reliability assessment and resilience, noting that the two fields are not mutually exclusive. Ryan considers how

Careful design of elicitation and analysis tools can lead to more insightful and reliable understanding of what happened in incidents and their causative factors. He notes the value of management support, training, feedback to staff and, perhaps most importantly, the presence of a 'no-blame' culture in capturing both incidents and 'near misses' and ensuring that knowledge and understanding of these feeds back into future work system management and design.

As noted in many other chapters in this volume, work is becoming increasingly influenced by technologies. Eason (Chapter 32) particularly considers the approach that should be adopted when introducing new technical systems. He highlights the positives and pitfalls from different methods of introducing new technology, ranging from a 'big bang' switch to phased implementation. In all of these approaches, however, it is of course key to engage with the workforce, ideally through participatory approaches (as considered in Chapter 34 of this book in Section VI) and to take a long-term perspective to understanding how the introduction of new technical systems will change work tasks and culture.

Finally in Section V, Sinclair and Siemieniuch (Chapter 33) present a philosophical discussion of systems ergonomics and associated methods. They particularly highlight the concept of 'systems of systems'. One of the consequences of a systems approach is that it increases the complexity of data collection and analysis; therefore, structured frameworks, such as presented in this chapter, to guide our thinking when adopting a systems approach, can be very valuable and help us to not only capture the context of work but also make clear and considered recommendations.

**Sarah Sharples**

---

# 28 Sociotechnical Design of Work Systems

*Patrick Waterson*

## CONTENTS

|                                                                                   |     |
|-----------------------------------------------------------------------------------|-----|
| Introduction.....                                                                 | 753 |
| Sociotechnical Systems Theory .....                                               | 754 |
| Background History .....                                                          | 754 |
| STS Constructs and Components .....                                               | 755 |
| Principles of Sociotechnical Design.....                                          | 756 |
| STS Models and Frameworks .....                                                   | 757 |
| Relation to Other Theories .....                                                  | 758 |
| Tools and Methods for STS Design and Analysis.....                                | 760 |
| Overview .....                                                                    | 760 |
| Methods Used to Analyse and Plan STS Design of Work Systems .....                 | 761 |
| Soft Systems Methodology .....                                                    | 761 |
| Scenario-Based Methods and Tools .....                                            | 763 |
| Methods Used to Evaluate STS (Re-)design of Work Systems .....                    | 764 |
| Checklists and Decision Criteria .....                                            | 764 |
| Other Methods.....                                                                | 765 |
| STS in Action: Three Case Study Examples.....                                     | 766 |
| Manufacturing and Shopfloor Jobs .....                                            | 766 |
| Manufacturing and ‘Virtual Teams’ .....                                           | 767 |
| Healthcare: Moving towards ‘Paperless’ Working in the Outpatient Department ..... | 767 |
| Future Challenges for STS.....                                                    | 768 |
| Need for New Theoretical Models .....                                             | 768 |
| Complex Systems and STS .....                                                     | 769 |
| Methods and the Scope of STS .....                                                | 769 |
| References.....                                                                   | 769 |

## INTRODUCTION

The term sociotechnical systems (STS) dates back to work carried out in the 1950s by a group of researchers at the London Tavistock Institute of Human Relations. The work of Eric Trist and Fred Emery initially focused on understanding the role of human skill and methods of working (e.g. team working) on productivity within coal mines (Trist and Bamforth, 1951; Emery and Trist, 1960). The primary motivation for the term sociotechnical system was to underscore the role of choice and organisational design in the interaction between people (the social systems) and tools, technologies and techniques (the technical system – Weisbord, 2012). A core value of the STS approach is that given the right choices, social and technical systems could be harmonised and balanced such that productivity, worker satisfaction and safety could be optimised in parallel (Cherns, 1976, 1987).

In more recent years, STS has influenced the development of a number of other different domains within ergonomics and human factors (E/HF). In particular, STS forms the basis of

macroergonomics and systems ergonomics. Macroergonomics is defined by Hendrick and Kleiner (2001) as 'a top-down sociotechnical system approach to the design of work systems and the application of the overall work-system design of the human-job, human-machine, and human-software interfaces'. Systems ergonomics emerged within the United Kingdom and Europe in the 1960s and builds on STS in viewing complex systems, for example, organisations, teams and types of technology, as composed of interrelated components, the properties of which are changed if the system is dissembled in any way (Katz and Kahn, 1966). The importance of adopting a systemic viewpoint within E/HF is underscored by Wilson (2012), who argues that studies need to examine not only individual (i.e. micro) issues but also wider social and organisational factors in parallel (i.e. macro issues) in order to be considered as E/HF investigations (see also Chapter 33 for further discussion of systems ergonomics methods).

Proponents of the STS approach argue that effective performance (defined in terms of production output, quality, employee well-being, job satisfaction) is a function of the degree to which design of the social and technical systems is considered to be complementary, as well as the relative effort placed upon the parallel design of both systems. If more effort is placed on the design of one system (e.g. the technical) compared to another, then suboptimal performance of the overall system is predicted to be the net outcome. The STS approach is often seen as an attempt to avoid technology-led approaches to work design (Blackler and Brown, 1986) and at the same time place emphasis upon the value and benefits of human-centred design. Biases towards technological, as compared to human concerns, are frequently viewed as a partial explanation for the underperformance of technology in the workplace (Landauer, 1997), as well as similar problems that apply to the use of new manufacturing practices (e.g. total quality management, business process re-engineering) within industry as a whole (Mumford and Hendricks, 1996; Waterson et al., 1999).

Sociotechnical studies of the workplace stretch back to the earliest days of research in E/HF in the immediate period following World War II. STS has also influenced many other areas, including large parts of organisational psychology, systems theory and industrial sociology (van Eijnatten, 1993). Within the field of E/HF, the list of applications of the STS approach is extensive, covering all of the aspects associated with work systems including the design of technology and jobs and arrangements for teamworking. Likewise, the range of work domains in which STS has been applied is also long. These include established areas such as human-computer interaction and occupational safety (e.g. Eason, 1988; Barling et al., 2002) to more recent forms of work involving knowledge management and electronic trading (e.g. Damodaran and Olphert, 2000; Coakes et al., 2002).

In this chapter, we will first look at the background history of the STS approach (alongside its main principles, constructs, components and models), before moving on to examine some STS methods and tools for the analysis and design of work systems. The next section describes a set of examples of application of STS to a variety of work systems (e.g. manufacturing, healthcare). A final section briefly describes some of the outstanding challenges for STS in the future and areas in need of further development.

## **SOCIOTECHNICAL SYSTEMS THEORY**

### **Background History**

In what has become something of a classic paper, Trist and Bamforth (1951) describe how workers had organised themselves into small, relatively autonomous work groups consisting of eight miners, who were responsible, as a group, for a full cycle in the coal extraction process. What was particularly interesting about this form of work organisation was that it was more efficient as compared to the prevailing method of working prescribed by management of the mine, which tended to emphasise a 'one best way' or Tayloristic (i.e. strict demarcation of roles and responsibilities with few opportunities for employee involvement) method of working.



Throughout the 1950s and 1960s, a number of other STS studies were conducted worldwide and in numerous contexts and work settings (e.g. the Indian textile industry, Rice, 1958; clothing factories in Norway, King, 1964). Most of these studies continued the line of Bamforth and the other Tavistock research workers in examining the value of work redesign as initiated and implemented by workers and pointed to the benefits of their participation in the design of work systems. In addition, the studies also demonstrated the value of forms of work organisation that were a mixture of formal and informal design and where work groups were given a degree of ownership and authority over the way in which work was carried out. These so-called semi-autonomous or autonomous work groups (AWGs) were viewed amongst followers of the STS approach as being one of the optimal ways of maintaining high levels of productivity and output and at the same time ensuring that the psychological needs of the individual worker were met (e.g. by providing a satisfying job and more generally a better quality of working life). In more recent years, AWDs and other similar forms of teamworking arrangements (e.g. high-performance work teams and flexible work groups – Kelly, 1982) have been the subject of a number of other evaluations. For example, studies have shown that this form of work organisation can bring about greater employee satisfaction (Trist et al., 1977), as well as reductions in operational costs due to employee innovation (Walton, 1977). Alongside recommendations concerning work groups and employee participation, the STS approach also makes use of a number of other principles, constructs and approaches to understanding work design.

### STS Constructs and Components

From the point of view of the STS approach, work design is seen as being influenced by many factors that depend upon the context in which it takes place; in other words, design is situation specific. There is therefore a need for a high degree of flexibility and choice in terms of what parts of the work system (e.g. individual jobs, managerial responsibilities, aspects of the technology being used) should be redesigned and precisely how this redesign should be done.

The STS approach also encompasses a set of values that apply to work design of all types (such as rich or lean technological environments, manufacturing or service industrial sectors) and involve all of its various components (i.e. technology, people, environment). Mumford and Axtell (2002), for example, describe how a great deal of emphasis was placed early on within the work of the Tavistock group, on the potential of STS interventions to improve the quality of working life of the individual worker. A central concern of the STS approach is that individuals involved with the work system should be treated with dignity and respect, and their aspirations and needs catered for. A similar concern applies within more recent Tavistock work (e.g. Heller, 1997), which emphasises the need to take into account the views of people when environmental change is being introduced.

One other important aspect of the STS approach that should be mentioned is the stress it places upon the value of participation and involvement in proposed changes to an existing work system or the design of a new one (see Chapter 34 for further discussion of participatory ergonomics). Proponents of the STS approach argue that the best way to achieve successful change is by involving all those likely to be affected by change in the development process. Participation in the change process is beneficial since it increases levels of ownership of the change process amongst employees, as well as reducing the likelihood that the proposed changes may be rejected by end users or other stakeholder groups. Amongst STS practitioners there is also awareness that participation has itself to be managed carefully and in many cases supported by tools and methods specifically designed to facilitate involvement and collaborative design (Greenbaum and Kyng, 1991; Wilson and Haines, 1997; and Chapter 34 in this book).

While there have been few critics of the underlying philosophy and motives of STS proponents and researchers, there have been criticisms aimed at the apparent lack of theoretical coherence within the approach (e.g. Kelly, 1978 – a theme revisited later on in the chapter). These criticisms have to some extent resulted in a number of attempts to systematically describe some core principles for STS.

## Principles of Sociotechnical Design

In a series of papers published in 1976 and 1987, the late Albert Cherns laid out what are generally recognised as some of the important principles of the STS approach (Table 28.1).

The principles take the form of a set of guidelines that can be used to help to design a safe and healthy work system and made up of jobs that are rewarding and satisfying for employees. The work system is typically made up of one or more persons, some form of technology in computerised (e.g. desktop computers) or non-computerised form (e.g. a factory production line), as well as environmental characteristics and influences surrounding the overall work system. For example, principle 2 (minimal critical specification) is an attempt to deal with the problems of overdesigning the work system so that, for example, individual roles and/or tasks to be performed by the technology (i.e. automated) are overly prescribed by the designer. Instead, Cherns suggests that work systems should be designed so that they are flexible, ‘minimally specified’ and where possible, designed partly by the individuals themselves within their work teams (principle 3). Other principles aim to improve the well-being and job satisfaction of individuals (e.g. principles 4, 7 and 8), as well as aspects of the management of team boundaries and communication requirements (principles 5 and 6).

More recently, there have been other attempts to provide principles and guidelines for socio-technical design. Clegg (2000), for example, has provided a set of principles that aim to update the earlier list from Cherns, as well as apply more directly to the domain of systems and technological design. Clegg has expanded the original nine principles into three different overlapping groups making a total of 19 principles in all. The first group of principles described by Clegg fall into the category of what he terms ‘meta-principles’, that is, principles which cover the philosophy or worldview underpinning design activity. The second and third groups cover the content and process of design, respectively. Table 28.2 outlines some examples of the three types of principles.

In some respects, the term principles is misleading since it might be taken to imply that the contents of Tables 28.1 and 28.2 should be applied in a mechanistic or prescriptive manner. However, as Clegg (2000, p. 474) points out, STS principles are not designed as blueprints that should be strictly adhered to, but rather as ‘ideas for debate, providing rhetorical devices through which detailed design discussions can be opened up and elaborated’. The principles are also designed to be used alongside methods and tools that support different types of sociotechnical activity, for example

**TABLE 28.1**  
**Examples of Cherns’ (1976, 1987) Principles of Sociotechnical Design**

| Principle                      | Summary                                                                                                             |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Compatibility                  | The process of design must be compatible with its objectives.                                                       |
| Minimal critical specification | No more detail in design than is needed, but design must express the essential requirements.                        |
| Sociotechnical criterion       | Control is local and should be given to the immediate work team – the aim is therefore to make supervision minimal. |
| Multi-function                 | Individuals and groups need a range of tasks to provide satisfying jobs and for redundancy and flexibility.         |
| Boundary location              | Boundaries are political boundaries and should be allocated managerial resources.                                   |
| Information flow               | Information should flow initially to the prime user or group and should avoid intermediaries where possible.        |
| Support congruence             | Systems should be established within a framework of social support for desired behaviour.                           |
| Design and human values        | Emphasis in design is placed upon the quality of working life.                                                      |
| Incompletion                   | Design is iterative and continuous.                                                                                 |

**TABLE 28.2****Examples of Clegg's (2000) Principles of Sociotechnical Systems Design**

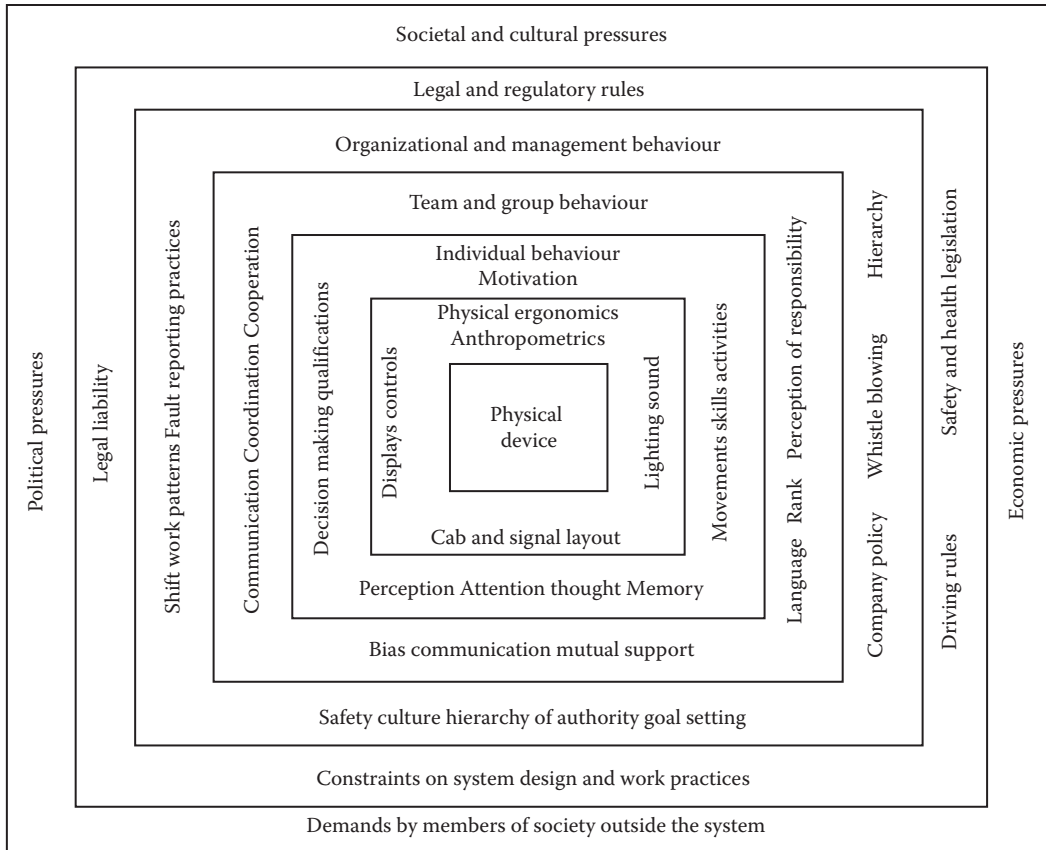
| Principle                                                                         | Summary                                                                                                                                                                                                                              |
|-----------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Meta-principles</i>                                                            |                                                                                                                                                                                                                                      |
| Design is systemic.                                                               | All aspects of system design are inter-connected. Leaving out one part (e.g. human aspects) will inevitably lead to suboptimal performance of the whole system.                                                                      |
| Values and mindsets are central to design.                                        | Attitudes and values on the part of designers, managers and/or users will shape the outcomes of design. It may be helpful at the outset of design to state clearly any possible biases and leanings towards possible design options. |
| Design is socially shaped.                                                        | Design is subject to social movements and trends; these may sometimes manifest themselves as fads and fashions.                                                                                                                      |
| <i>Content principles</i>                                                         |                                                                                                                                                                                                                                      |
| Systems should be simple in design and make problems visible.                     | Design simplicity includes ease of system use, understanding and learnability.                                                                                                                                                       |
| Design entails multiple task allocations between and amongst humans and machines. | Task allocations should be systematically analysed and evaluated. In the case of deciding to allocate tasks to a team or AWG, then decision criteria should be used (e.g. Medsker and Campion, 1997).                                |
| Problems should be controlled at source.                                          | Problems should be tackled by those closest to the problem since this has motivational and logistical advantages (e.g. people like to have control over the problems they face).                                                     |
| <i>Process principles</i>                                                         |                                                                                                                                                                                                                                      |
| Evaluation is an essential part of design.                                        | Evaluation, although rarely undertaken, has several advantages, the chief one being that an organisation can learn from its mistakes and successes.                                                                                  |
| Design involves multidisciplinary education.                                      | There is a need for a diverse range of expertise and skills within design in order to bring about innovation, as well as viewing design from several perspectives.                                                                   |
| Resources and support are required for design.                                    | Successful design comes at a price and in order to follow some of the principles, there will be a need for organisations to invest resources into the process.                                                                       |

forming an overview of a particular design or carrying out the evaluation of design options. Some examples of these tools are described in the next section.

## STS MODELS AND FRAMEWORKS

One of the most common ways of describing STS is to use a framework or model to describe the various components within work systems and then to outline the human factors issues which apply at each level. In some cases, the models can be used to carry out specific tasks or goals (e.g. accident analysis and investigation – Salmon et al., 2011); in other cases, they may be used to capture some of the STS properties of as they relate to a specific application domain (e.g. transportation, health-care). Figure 28.1 shows an example of a model (Moray's 1997 'onion' model) which was developed within the context of railway human factors and safety. The model describes a number of levels of analysis (societal and cultural pressures) through to aspects of the physical device or technology (individual behaviour). Each level is associated with a set of HF issues such as communication (team and groups behaviour), lighting and sounds (individual behaviour) and safety culture (organisational and management behaviour).

Models have also been developed for other domains – Figure 28.2 shows an example of a systems model for health information technologies (Carayon et al., 2006). The model aims to understand the role of individual, organisational, task-based and technological factors in parallel. These factors in combination impact upon the way in which care processes are delivered and outcomes for patients such as quality of care and safety.

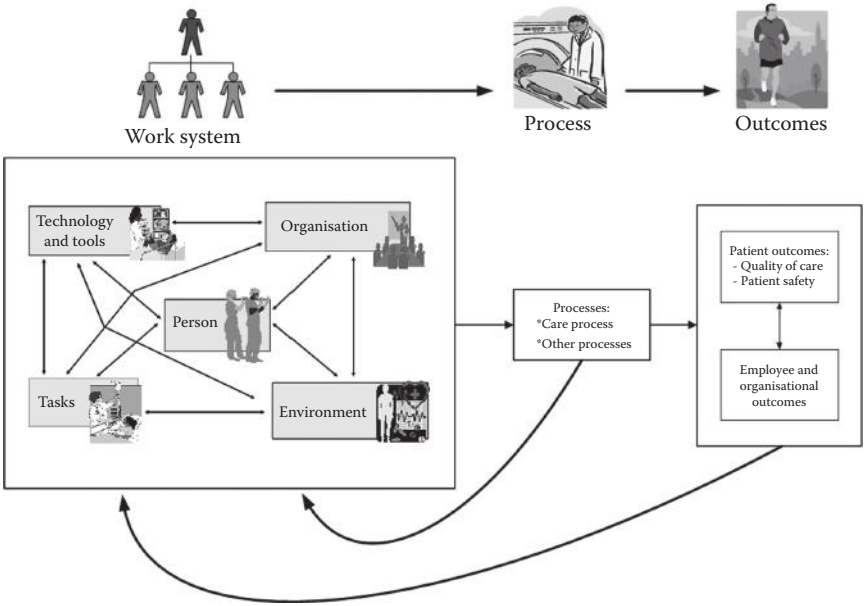


**FIGURE 28.1** STS 'onion' model. (From Moray, N., Real prediction of real performance, in: Wilson, J.R., Norris, B., Clarke, T., and Mills, A. (eds.), *People and Rail Systems: Human Factors at the Heart of the Railway*, Ashgate, London, U.K., 2007.)

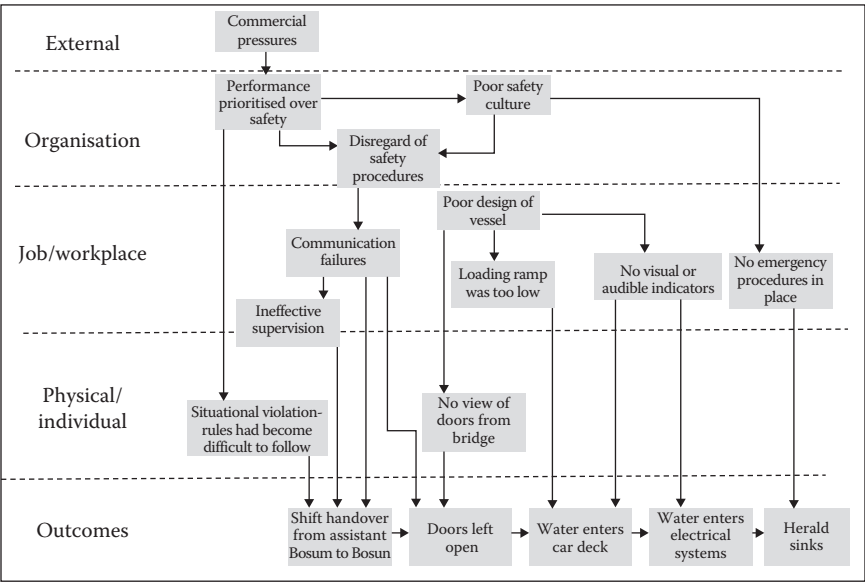
A final model aims to capture some of the factors which contribute to the occurrence of major accidents and catastrophes involving complex sociotechnical systems. One of the main drivers for developing these models is to shift emphasis away from viewing accidents as the result of a single cause or the actions of a specific individual (the so-called 'bad apple' theory of accident causation – Dekker, 2006). From this point of view, accidents arise as a result of a complex set of events and circumstances which involve multiple system components. The AcciMap model, for example (Rasmussen, 1997; Rasmussen and Svedung, 2000), organises these components using a set of headings ranging from influences which are external (e.g. political or regulatory influences), organisational (e.g. safety culture) to physical/individual (e.g. equipment failures). Figure 28.3 shows an example of an AcciMap which describes some of the events leading up to the sinking of the Herald of Free Enterprise passenger ferry in Zeebrugge in 1987. AcciMaps are also related to the concept of system 'resilience', which is discussed in more detail later on in this chapter.

### Relation to Other Theories

In recent years, STS has been linked to a number of other theories and approaches. One example of such a theory is adaptive structuration theory (AST). AST is often used to understand change processes which occur when information systems are implemented and adopted; this is achieved by examining two vantage points in parallel: (1) the types of structures that are provided by



**FIGURE 28.2** The Systems Engineering Initiative for Patient Safety (SEIPS) model. (From Carayon, P. et al., *Qual. Saf. Health Care*, 15(Suppl. 1), i50, 2006.)



**FIGURE 28.3** Herald of Free Enterprise accident AcciMap. (Adapted from Rasmussen, J. and Svedung, I., *Proactive Risk Management in a Dynamic Society*, Swedish Rescue Services Agency, Karlstad, Sweden, 2000.)

information systems (e.g. the functionality of the technology) and (2) the structures that emerge in human action as a result of people interacting with the new technology (e.g. patterns of technology use and adoption). AST is based on structuration theory which was originally developed by the sociologist Anthony Giddens (1984). Structuration theory proposes that a complete understanding of social interaction requires incorporating explanations of both the structure of relationships and

the dynamics or processes of relationships. The processes both take place in social structures, and shape those structures over time. At the heart of AST is the role of advanced information technology and its appropriation by members of the organisation as they work together. The theory describes how a technology's inherent structural characteristics shape interaction patterns without determining the interaction in a definitive way. How people choose to adopt or appropriate the technology helps to shape their decision processes. Patterns of social structure emerge in the group over time and are influenced by use of the new technology. DeSanctis and Poole (1994) also suggest that analysis based on AST can identify certain patterns that are more likely to be associated with organisational effectiveness than others. Examples of combining STS and AST together are described in later sections of this chapter.

## **TOOLS AND METHODS FOR STS DESIGN AND ANALYSIS**

### **Overview**

As mentioned earlier, a characteristic of the STS approach is its flexibility. One consequence of this is that there are no hard and fast rules or guidelines regarding what types of methods should be used in a study or where and when these should be applied. The types of methods to be used are largely up to the researcher and may be determined by factors such as the specific nature of the problem to be addressed, the degree of access granted the researcher (e.g. access to company personnel, documentation and other data), the timescale of the study, the views and requirements of study participant regarding the expected outcome from the study, and the type of feedback that is to be given to the participants involved in the study. This latter point is especially important since the STS approach often involves what is known as 'action research' (Foster, 1972).

Action research typically involves the STS researcher adopting the role of a participant in the change process (as compared to a more traditional, academic approach which is often more removed from the process). This role also involves the commitment that the researcher will seek practical solutions to the problem being addressed, as well as examining more theoretical issues. For example, a researcher involved in the study of the introduction of a new IT system within a large company might try to help the company find ways of involving users in the change process (e.g. by helping to run user workshops), while at the same time seeking to address research questions such as how the technology changes or restructures patterns of working in the light of theories within the scientific literature.

As well as involving a practical objective, the STS approach also typically employs more than one method of study when carrying out a research project. Wall et al. (1986), for example, carried out semi-structured interviews during the course of their study of the introduction of AWGs within a company manufacturing sweets while at the same time asking shopfloor workers to complete a questionnaire survey during the early and later phases of the change process (a so-called quasi-experimental field study – see Cook et al., 1990 for further details of this research method as well as Chapter 1 of this book for discussion of such approaches). The important point to note is that STS studies often employ a variety of methods and these are not exclusively limited to either quantitative or qualitative approaches. Furthermore, application of the STS approach may also involve many of the methods described in other chapters in this volume (e.g. Chapters 6 and 7).

The next sections review methods which are slightly less well known and do not easily fit within the quantitative/qualitative distinction between research methods; nevertheless, they are often used within STS studies. A rough distinction can be made between methods that support the analysis and planning of design based upon STS (e.g. soft systems methodology and the use of scenarios) and those that are intended to provide help in evaluating the outcomes of design and make it possible to suggest areas of the work system that can be redesigned (e.g. checklists and decision criteria).





**TABLE 28.3****Stages of Applying SSM to the Analysis of the Implementation of the CAD/CAM Tools within the Engineering Organisation**

| Stage | Description                             | Summary of Activities and Background                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------|-----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1.    | Problem situation                       | <p>Aerospace engineering company employing 2000 people.</p> <p>Complex organisational structure, three major functions: design, manufacture and corporate engineering.</p> <p>Design and manufacture functions operate autonomously; in the past there has been little success in integrating the two functions.</p> <p>Senior managers decide to invest in CAD/CAM in order to reduce costs and improve quality and improve integration between design and manufacturing.</p> <p>Project team is set up to manage implementation; researchers are asked to investigate the implementation process and make recommendations for improvements from an organisational (as compared to technical) point of view.</p> |
| 2.    | The 'rich picture'                      | <p>Research carried out over a period of 18 months.</p> <p>Variety of data collected including interviews (n = 35), periods of observation and attendance at meetings and questionnaire survey (n = 36).</p> <p>The main concern of the staff at headquarters was getting the technology implemented at a minimum cost.</p> <p>Little attention was paid to the wider organisational impact of CAD/CAM or the opportunities the implementation provided for reorganising work in the two functions (i.e. design and manufacturing).</p>                                                                                                                                                                           |
| 3.    | Relevant systems and 'root' definitions | <p>The researchers decided to describe both the existing relevant systems that they believed were guiding the implementation, as well as an alternative system and its root definition.</p> <p>The researchers argued that the project was seen and managed as a technology implementation project (as compared to an opportunity for organisational change).</p>                                                                                                                                                                                                                                                                                                                                                 |
| 4.    | Conceptual model                        | <p>In order to meet the demands of the alternative root definition (i.e. the implementation as an organisational change), the researchers developed a conceptual model.</p> <p>The model consisted of seven elements in a flowchart; these were the following: theorise → resource → plan → participate → educate → support → evaluate.</p> <p>The new model was used as the basis for a set of recommendations to the company.</p>                                                                                                                                                                                                                                                                               |
| 5.    | Comparison and agenda                   | <p>Comparison between the new model and the real events going on in the company resulted in the identification of some major gaps.</p> <p>These gaps led the researchers to generate an agenda for change.</p> <p>Some of the key items in the agenda included investing more resources in the change programme, encouraging more participation by end users and placing more emphasis on wider education, awareness and communication for end users and others.</p>                                                                                                                                                                                                                                              |
| 6.    | Debate                                  | <p>An open report describing the finding, root definitions and agendas for change was presented to all interested parties within the companies.</p> <p>End users expressed satisfaction with the outcomes and acknowledged the technology-led aspect of the implementation.</p> <p>Senior managers, however, were less convinced and disagreed with many of the conclusions from the report.</p>                                                                                                                                                                                                                                                                                                                  |
| 7.    | Implementation                          | <p>Some of the recommendations from the report were implemented.</p> <p>In particular, a CAD/CAM user group was set up in order to address issues related to end user participation, education and training, for example.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

Sources: Symon, G.J. and Clegg, C.W., *J. Occup. Psychol.*, 64, 272, 1991a; Symon, G.J. and Clegg, C.W., Implementation of a CAD/CAM system: The management of change at EML, in: Legge, K., Clegg, C.W., and Kemp, N.J., (eds.), *Case Studies in Information Technology, People and Organisations*, Oxford, U.K., NCCC Blackwell.

**TABLE 28.4**  
**Elements in the Root Definition of the CAD/CAM Implementation**

|                                                                                                   | Existing System                                                                                                | Alternative System                                                                                                                                |
|---------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>C</b> – Customers (or clients) – i.e. the people who receive what the system does              | Corporate engineers and designers in the different product groups                                              | Same                                                                                                                                              |
| <b>A</b> – Actors – i.e. the people who carry out what the system does                            | Corporate engineering (senior managers)                                                                        | Everyone involved with the change (e.g. managers, end users) – i.e. a participative approach to the implementation                                |
| <b>T</b> – Transformation – i.e. what the system changes, from one state to another               | Replacing a manual system for design and manufacture with a CAD/CAM system                                     | Designing a new organisational system, integrating design and manufacturing and assisted by the CAD/CAM system                                    |
| <b>W</b> – Weltanschauung – i.e. the underlying worldview                                         | Emphasis is upon technical issues and problems, the view being that organisational issues can be handled later | Joint (i.e. sociotechnical) consideration of technical, strategic and organisational issues                                                       |
| <b>O</b> – Owners – i.e. those who have power over the system, e.g. to cause it to cease to exist | Corporate engineering (senior managers)                                                                        | Everyone involved with the change (e.g. managers, end users) – i.e. a participative approach to the implementation                                |
| <b>E</b> – Environmental – constraints – i.e. constraints that the system has to take as given    | A minimum of scarce company resources (i.e. time and money)                                                    | Process will require appropriate resources to operate (in some cases more money will have to be invested in addressing the organisational issues) |

SSM can prove to be a very useful method for gathering a wide variety of opinions regarding a new system. The method has several strengths, including the fact that it helps to include a wider set of perspectives (e.g. by involving stakeholders in the analysis and implementation of change) than is perhaps normally the case, in the overall process of managing change. Secondly, SSM provides a systematic means of organising complex data and representing it in a simple and concise manner. Finally, the methodology encourages a participative and systemic approach to design (i.e. it is congruent with some of the sociotechnical principles described earlier in Tables 28.1 and 28.2).

Needless to say, SSM also has some disadvantages (Clegg and Walsh, 1998, pp. 231–232), these include the fact that in many cases the language and terminology used by SSM (e.g. ‘rich picture’, ‘root definition’) can be complex to use and describe. Furthermore, SSM also offers little guidance in the details of how to implement some of the outputs from its use; perhaps for this reason it is best used in combination with some of the other methods described in this section rather than in isolation.

### Scenario-Based Methods and Tools

One of the most common techniques used by STS practitioners, as well as others drawn from a variety of backgrounds (e.g. human factors, software engineering, human-computer interaction), are scenarios. Carroll (2000) defines a scenario as an ‘informal narrative description’, and the term is widely used to convey some kind of story-like representation that can serve as the basis for discussion and debate amongst groups such as system end users and other stakeholders. One of the main advantages of scenarios is that they are a natural way for people to explain what they are doing or how something can be achieved (Preece et al., 2002). Within the context of generating the requirements for a new system or a new set of work-based roles, for example, scenarios offer a simple way

**TABLE 28.5****Summary of the Scenarios Generated Using the Tool in the Workshops**

|           | <b>Scenario 1</b>                                                                                   | <b>Scenario 2</b>                                                                                | <b>Scenario 3</b>                                                                                               |
|-----------|-----------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| Scope     | Existing structure                                                                                  | Partial multi-skilling                                                                           | Complete multi-skilling                                                                                         |
| Structure | Maintain present boundaries: solutions personnel work upstairs<br>Melting personnel work downstairs | Maintain boundaries but melters share some tasks with solutions personnel                        | No boundaries. Abandon distinction between solutions and melting personnel<br>Work as one crew and rotate tasks |
| Roles     | Specialised and separated                                                                           | Overlap of tasks. Melters to become partly multiskilled<br>Solutions to carry out existing tasks | Single role for all crew members combining all tasks                                                            |

*Source:* Nadin, S.J. et al., *Hum. Factors Ergon. Manuf.*, 11(1), 53, 2001.

of describing the existing way in which the work system is operating, as well as planning the design of a new system that is sensitive to the needs of people working within it.

Nadin et al. (2001) used scenarios to help redesign a set of jobs and task carried out by shopfloor workers in a company producing photographic film and paper. At the time of the study, the company was planning to introduce a new computer-based scheduling system, and the likelihood was that the new system would change the work of two main types of jobs ('solutions' and 'melting' personnel) within its chemical processing department (e.g. changing the tasks they performed and the way in which they interacted). One part of the study involved generating a set of scenarios that described the existing way in which work was carried out by the two job types, as well as alternative involving different levels of skill distribution and interaction between the two types of personnel. The scenarios were generated during a series of three workshops involving shopfloor workers and managers and were structured around three main headings (the scope and the structure of the scenario and the roles involved) – see Table 28.5.

The scenarios proved to be popular amongst the participants at the workshops largely because they helped to structure discussion centred on the change process, as well as having the added benefit that all of those likely to be affected by the change were involved in generating the scenarios. In many cases, the use of scenarios can prove worthwhile since they help participants to consider issues that would have been otherwise ignored or overlooked. This is especially important given the stress placed upon joint consideration of both social and technical aspects of change management, as well as other systemic issues (see principle 1 in Table 28.2), within the STS approach.

## **METHODS USED TO EVALUATE STS (RE-)DESIGN OF WORK SYSTEMS**

### **Checklists and Decision Criteria**

At the heart of the STS approach is a concern that the design of work systems should go beyond the planning and generation of design options, but, critically, should also involve their systematic evaluation and subsequent refinement (c.f. the principles in Tables 28.1 and 28.2). A number of STS evaluation methods exist and some of these are aimed at covering the larger work system (e.g. the organisation as a whole or a department within the organisation – see Hendrick, 1997 for examples) or specific work sub-systems (e.g. interfaces between people and technology – Mumford, 1995). Furthermore, some evaluation methods focus on specific areas or domains of design activity in more detail (e.g. usability evaluation, Nielson, 1997; requirements engineering, Eason et al., 1996). Two of the most common evaluation methods that can be used to evaluate sociotechnical aspects of the whole work system, as well as its parts, are the use of checklists and decision criteria.

**TABLE 28.6****Example Questions in the Form of a Checklist for the Design of Teams**

1. Are worker's tasks highly interdependent or could they be made to be so? Would this interdependence enhance efficiency or quality?
2. Do the tasks require a variety of knowledge, skills and abilities such that combining individuals with different backgrounds would make a difference in performance?
3. Is cross-training desired? Would breadth of skills and workforce flexibility be essential to the organisation?
4. Could increased motivation and effort to perform make a difference in effectiveness?
5. Can social support help workers deal with job stresses?
6. Could increased communication and information exchange improve performance rather than interfere?
7. Could increased cooperation aid performance?
8. Are individual evaluation and rewards difficult or impossible to make or do workers mistrust them?
9. Could common measures of performance be developed and used?
10. Is it technically possible to group tasks in a meaningful, efficient way?

*Source:* Adapted from Campion, M.A. and Medsker, G.J., Job design, in: Salvendy, G. (ed.), *Handbook of Industrial Engineering*, John Wiley & Sons, New York, 2001.

Checklists can be used by researchers, as well as others outside of the E/HF profession (e.g. managers), to assess the degree to which an existing or a planned work design compares to established measures, or indicators, of good practice. Campion and Medsker (1992) and Medsker and Campion (1997), for example, describe a checklist that can be used to assess the degree to which teams have been successfully implemented within an organisation (Table 28.6).

Checklists can also be used to benchmark the progress of changes to a work system as they occur over time. For example, they can sometimes be turned into other material, such as survey questionnaires, and used as an important means of gaining feedback from shopfloor workers and other personnel involved in the change process. It should also be noted that these types of methods are often more successful when they are accompanied by further guidance covering strategies and other advice that can be used to implement changes to the work system (Parker and Wall, 1998; Clegg, 2000). A final set of methods that can be used to evaluate options for work design are the various forms of decision criteria that exist for STS evaluation.

Methods that use decision criteria typically involve a group of stakeholders involved with the design process collectively evaluating design options using a pre-defined categorised list of issues that are known to be of relevance within the design domain. The advantage of the pre-defined list is that aspects of the design of the system that might either be overlooked or ignored can be systematically addressed and evaluated. Table 28.7 outlines a categorised list of decision criteria that were developed for use within the domain of task allocation, but can also be used for wider STS work design (Waterson et al., 2002).

A typical situation in which the decision criteria in Table 28.7 are used is where there are a number of design options for a particular set of work tasks. Stakeholders work through the criteria and consider the design options in the light of questions that are raised as a result of consulting the criteria. For example, stakeholders might examine the category 'Cultural/environmental issues' and pose the question – 'Do any of the design options have cultural/environmental concerns?' In order to provide stakeholders with some help, the category is then broken down into criteria that should be taken into account and posed as sub-questions (e.g. 'Do any of the design options raise data security issues?').

## OTHER METHODS

Aside from the methods we have described, there exist a number of others that can be used for STS analysis, planning and evaluation. Many of these have much in common with SSM, scenarios and

**TABLE 28.7**  
**Examples of Decision Criteria**

| Category                            | Criteria                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Goal issues                         | Goals of the system, goals of the organisation, other goal issues                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Organisational issues               | Existing/required systems and procedures, existing/required organisational norms and practices, organisational requirements, organisation structures, accountability, other organisation issues                                                                                                                                                                                                                                                                                                                                                          |
| Cultural/environmental issues       | Cultural issues, legal requirements, health and safety, data security, national/international political considerations, other cultural/environmental issues                                                                                                                                                                                                                                                                                                                                                                                              |
| Resource issues                     | Technology; people; money; other investments (e.g. time); knowledge, skills and education; other resource issues                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| People issues                       | Level, type and number; knowledge, skills and education; trust; training; social acceptability; other people issues                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Task issues                         | Speed of response, accuracy, frequency, physical demands, cognitive demands, emotional demands, reliability, efficiency, flexibility, operational criticality (likelihood and implications), safety criticality, uncertainty of occurrence, uncertainty of the 'type of problem', uncertainty of the 'correct response', variability in performance, task interdependencies, redundancy, other task issues                                                                                                                                               |
| Job design/work organisation issues | Control (e.g. autonomy, responsibility, variance handling, decision making), accountability, variety, skill use and development, workload, clarity of goals and requirements, feedback on performance, communication and social contact, health and safety (mental and physical), flexibility, information exchange and knowledge sharing, situation awareness, understanding of the system, mutual adjustment/team collaboration, role ambiguity and overload, motivation, satisfaction, stress, performance, other job design/work organisation issues |
| Technology issues                   | Feasibility of automation; cost of automation; maintainability; reliability; level of performance; trust in technology; level, type and amount of technology; other technology issues                                                                                                                                                                                                                                                                                                                                                                    |

Source: Waterson, P.E. et al., *Hum. Factors*, 44(3), 376, 2002.

the use of decision criteria. Some of the most important include the group of methods that can be subsumed under the title of 'stakeholder analysis' (Burgoyne, 1994), of which SSM is a close relative. In addition, in the last few years, methods that encourage and support greater user involvement in the design process (e.g. 'rapid prototyping', Mayhew, 1999; Scandinavian participatory design, Bodker et al., 1991) have become popular. Likewise, a technique that is similar in its analytical aims to that of tracer studies, known as 'variance analysis' (Davis and Wacker, 1982), is often suggested as an effective means of designing work systems that utilises the STS approach.

## STS IN ACTION: THREE CASE STUDY EXAMPLES

This section describes three case studies drawn from the domains of manufacturing and healthcare. Work on manufacturing is often seen as one of the 'home territories' of STS research and practice, particularly as it relates to job and work design. Accordingly, the first case study examines job design changes as they took place on the shopfloor. Two other case studies describe technological developments as they relate to more recent developments, namely, the introduction of virtual teams within organisations, and drives to move towards paperless, electronic-supported working.

### MANUFACTURING AND SHOPFLOOR JOBS

Buchanan and Boddy (1983) describe an example of a case study which used the STS approach to evaluate the impact of a new computer system upon the work of biscuit-making operators. They found that different types of application of the new computer technology had very different effects



upon the work of operators within the company. The system which controlled dough-mixing tasks largely replaced the craft-based skills of doughmen and led to simplified jobs which were less satisfying than those before the introduction of the system. By contrast, the part of the system which controlled weighing tasks largely complemented the work of other operators and led to more interesting and rewarding jobs. The differential impact of the new technology could largely be explained by the objectives of managers and the way in which the new system was planned and introduced. In the case of mixing tasks, the system largely removed a number of tasks which were viewed as being important in terms of control and decision-making responsibilities amongst doughmen. These tasks were automated, and managers failed to consider the impact such automation would have upon the quality of life and overall job satisfaction of operators. On the other hand, the new weighing system retained a number of key tasks for operators and removed some tasks which were viewed as being repetitive or tedious to complete. For example, the new weighing system allowed operators to control weighing tasks more effectively and gain rapid feedback on their performance, leading to a more interesting and challenging job as a result. Overall, the case study demonstrated that the implementation of new technology needs to be sensitive to the views of those involved in the change and that the design of the technology should complement, rather than conflict with, existing skills and knowledge levels. It is therefore important that work is organised to develop the human skills and motivation required for overall system effectiveness.

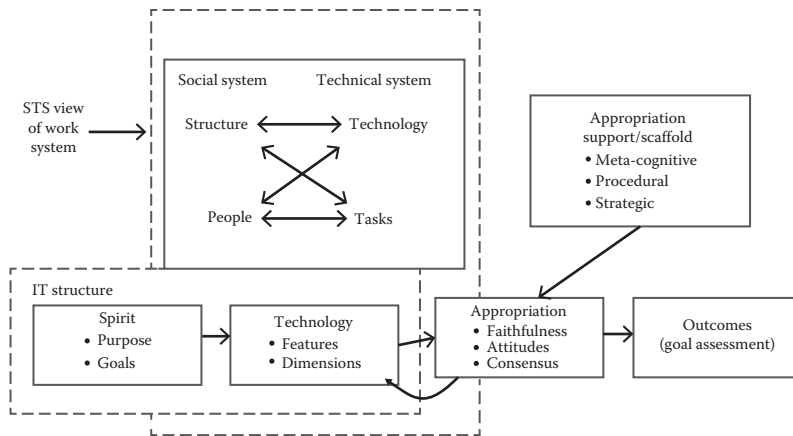
### **MANUFACTURING AND ‘VIRTUAL TEAMS’**

Maznevski and Chudoba (2000) report a longitudinal study of a large U.S. manufacturing company which made use of virtual teamworking to coordinate working arrangements across the United States and Europe. Over a period of 21 months the study examined communication patterns between members of three virtual teams and how these changed over time. The work of the teams involved a mixture of face-to-face interaction, as well as the use of communication technologies such as video conferencing, teleconferencing, email and telephone. In order to categorise how communication patterns evolved over time, Maznevski and Chudoba (2000) used a coding framework for their interviews and observations based on AST (as described earlier). The framework included codes covering four main aspects of the work of the virtual teams: (1) the structural elements which shaped the work of the teams (e.g. task characteristics such as the degree of interdependence requires by a task across team members), (2) technology appropriation (e.g. how the teams used communication technologies and which they chose to convey a specific message), (3) decision-making processes (e.g. idea construction, problem solving), and (4) decision outcomes (e.g. how teams rated the quality of decision making and the extent to which they enjoyed working together).

One of the most interesting findings from the study was that the most effective virtual teams established regular patterns of interaction which repeated over time. These temporal patterns typically involved an initial co-located, face-to-face team meeting with attempts by team members to resolve problems and build relationships. This ‘coordination’ meeting served like a ‘heartbeat’ which revitalised the relationships between team members and improve other aspects of team processes. Regular meetings of this kind helped to ensure that when the teams were working across distance and using technologies such as video conferencing, their work was cohesive and collaborative. One of the implications from the study is that virtual teams need to punctuate remote working with face-to-face meetings in order to ensure that the teamwork remains effective and productive.

### **HEALTHCARE: MOVING TOWARDS ‘PAPERLESS’ WORKING IN THE OUTPATIENT DEPARTMENT**

Waterson et al. (2012) and Waterson (2014) describe a case study which combined elements of STS and AST in order to understand the preparations which were being made for the introduction of ‘paperless’ working within a large U.K. hospital. The study took place in an outpatient department and focused on how the current work system for patient records worked, as well as the expectations



**FIGURE 28.5** Combined model with elements drawn from sociotechnical and structuration theory. (From Waterson, P.E. et al., *Int. J. Med. Inform.*, 81, 114, 2012.)

of staff and the likely outcomes of introducing technology in the form of an electronic patient records (EPR) system. The study used a combined STS and AST model (Figure 28.5) as a template to code interviews with staff within the outpatient department. The study found that one of the main influences on attitudes towards the type of EPR that should exist within the department was prompted by negative reactions to the way in which previous IT systems had been managed and attempted to be introduced in the past. A strong commitment to end-user involvement in the design of the new EPR system also shaped attitudes towards future expectations of the adoption of new EPRs within the department. In combination, these attitudes served as a 'scaffold' which influenced their 'mental model' of how the new EPR system might work. In addition, part of this appropriation scaffold involved a high degree of scepticism that a rapid change to 'paperless' working was likely to be possible.

## FUTURE CHALLENGES FOR STS

There are at least three major challenges facing researchers and practitioners within the area of STS in the future: the need to develop new theoretical models and draw on other theories of technology use; complex systems and STS; methods and the practical coverage of STS.

## NEED FOR NEW THEORETICAL MODELS

Much of the theoretical work within STS was developed in the 1950s and 1960s, and many of the core constructs and components (e.g. optimisation of social and technical systems, the concepts of redundancy and optimisation) need to be reappraised in the light of new theories and concepts. Much of the STS approach concerns itself with technological work, and it is not clear how STS theory compares with more recent attempts to understand the overlap between social, organisational and technological issues in these types of work domains (e.g. actor/network theory, Latour, 1990, 1992, 1999; distributed cognition, Hutchins, 1995). Some attempts have been made to update the original work associated with the Tavistock tradition of STS (e.g. Trist and Bamforth, 1951). The healthcare case study described earlier represents an attempt to integrate constructs from structuration theory within an STS framework, for example. Likewise, Walker et al. (2010) and Griffin et al. (2010) also integrate elements of STS with elements of distributed cognition and network analysis and apply this to air traffic control and accident causation. STS has much in common with other theories and further efforts to render STS theory and practice more coherent (see similar comments made by Kelly, 1978, over 30 years ago).

A second related theoretical challenge is to understand causality across levels within the overall sociotechnical system. Karsh (2003), for example, used the term mesoergonomics in order to describe an STS approach to ergonomic theory and research whereby the relationship between variables in at least two different system levels or echelons is studied, where the dependent variables are human factors and ergonomic constructs. Within a number of contexts and domains (e.g. safety culture, Zohar and Luria, 2005; patient safety, Karsh and Brown, 2010), it has been demonstrated that certain phenomena (e.g. errors and safety violations) occur because of the influence of a number of levels within the system (e.g. group and individual behaviour), and studying one system level in isolation runs the risk of designing interventions (i.e. improvements) which only apply to one system level and are therefore likely to be ineffective. Further work in this area is being undertaken in this area by Karsh et al. (2014); however, there is a need for more studies to adopt multilevel analysis and examine causality across and between system levels. Moreover, a holistic and pluralistic stance towards human factors and ergonomics is argued by some to be one of the defining characteristics of the discipline (Wilson, 2012).

### COMPLEX SYSTEMS AND STS

A second challenge for the future will be dealing with the increasing complexity and intractability of sociotechnical systems. Recent disasters such as the *Deepwater Horizon* explosion and the Fukushima Nuclear disaster have demonstrated the need to build what Hollnagel et al. (2006) refers to as 'resilience' into these systems in order to prevent their reoccurrence. Resilience is sometimes defined as the intrinsic ability of an STS to adjust its functioning prior to, during or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions (Hollnagel, 2012). How this can be achieved and how to trade off issues such as safety, reliability, costs and productivity against one another remains an open question for the future (Wilson et al., 2009). In the present environment, where governments and organisations face huge pressure to reduce costs and curtail budgets, these concerns are likely to take on more importance.

### METHODS AND THE SCOPE OF STS

A final consideration for the future is the need to provide more practical advice and support for practitioners who want to use STS to improve the workplace. Much of the debate within STS circles is conducted by academics and, as perhaps this chapter has indicated, much of it is concerned with theoretical issues. A huge variety of methods, techniques and tools exist; however, many of them only apply to specific application domains (e.g. requirements analysis). The extent to which these are used by human factors practitioners is questionable. Many of the methods prove to be difficult to use and are time-consuming. Moreover, many of the most practically oriented sets of tools developed in the past emerged out of work within job and work design (a topic which unfortunately seems to have declined in popularity amongst E/HF researchers in the last few years).

### REFERENCES

- Atkinson, C.J. (2000), Sociotechnical and soft approaches to information requirements elicitation in the post methodology era. *Requirements Engineering*, 5, 67–73.
- Barling, J., Kelloway, E.K. and Zacharatos, A. (2002), Occupational safety. In P.B. Warr (Ed.), *Psychology at Work* (5th Edn.). Harmondsworth, U.K.: Penguin Books.
- Bell, S. and Morse, S. (2013), How people use rich pictures to help them think and act. *Systemic Practice and Action Research*, 26, 331–348.
- Blackler, F. and Brown, C. (1986), Alternative models to guide the design and introduction of new technologies into work organisations. *Journal of Occupational Psychology*, 41, 211–221.
- Bodker, S., Greenbaum, J. and Kyng, M. (1991), Setting the stage for design as action. In J. Greenbaum and M. Kyng (Eds.), *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, NJ: LEA Press.

- Buchanan, D.A. and Boddy, D. (1983), Advanced technology and the quality of working life: The effects of computerized controls on biscuit-making operators. *Journal of Occupational and Organizational Psychology*, 56(2), 109–119.
- Burgoyne, J.G. (1994), Stakeholder analysis. In C. Cassell and G. Symon (Eds.), *Qualitative Methods and Analysis in Organizational Research*. London, U.K.: Sage Publications.
- Campion, M.A. and Medsker, G.J. (1992), Job design. In G. Salvendy (Ed.), *Handbook of Industrial Engineering*. New York: John Wiley & Sons.
- Carayon, P., Hundt, A.S., Karsh, B.-T., Gurses, A.P., Alvarado, C.J., Smith, M. and Brennan, P. (2006), Work system design for patient safety: The SEIPS model. *Quality and Safety in Health Care*, 15(Suppl. 1), i50–i58.
- Carroll, J.M. (2000), *Scenario-Based Design: Envisioning Work and Technology in System Development*. New York: John Wiley.
- Checkland, P. and Scholes, J. (1999), *Soft Systems Methodology in Action: A 30 Year Retrospective*. Chichester, U.K.: John Wiley & Sons.
- Cherns, A. (1976), The principles of sociotechnical design. *Human Relations*, 29, 783–792.
- Cherns, A. (1987), Principles of sociotechnical design revisited. *Human Relations*, 40, 153–162.
- Clegg, C.W. (2000), Sociotechnical principles for system design. *Applied Ergonomics*, 31, 463–477.
- Clegg, C.W. and Walsh, S. (1998), Soft systems analysis. In G. Symon and C. Cassell (Eds.), *Qualitative Methods and Analysis in Organizational Research*. London, U.K.: Sage Publications.
- Coakes, E., Willis, D. and Clarke, S. (Eds.) (2002), *Knowledge Management in the Sociotechnical World*. London, U.K.: Springer Verlag.
- Cook, T.D., Campbell, D.T. and Peracchio, L. (1990), Quasi-experimentation. In M.D. Dunnette and L.M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology*. Palo Alto, CA: Consulting Psychologists.
- Damodaran, L. and Olphert, W. (2000), Barriers and facilitators to the use of knowledge management systems. *Behaviour and Information Technology*, 16(6), 405–413.
- Davis, L.E. and Wacker, G.L. (1982), Job design. In G. Salvendy (Ed.), *Handbook of Human Factors*. New York: John Wiley & Sons.
- Dekker, S. (2006), *The Field Guide to Understanding Human Error*. Farnham, U.K.: Ashgate.
- DeSanctis, G. and Poole, M.S. (1994), Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, 5, 121–147.
- Eason, K.D. (1988), *Information Technology and Organisational Change*. London, U.K.: Taylor & Francis Group.
- Eason, K.D., Harker, S.D.P. and Olphert, C.W. (1996), Representing socio-technical systems options in the development of new forms of work organization. *European Journal of Work and Organizational Psychology*, 5(3), 399–420.
- Emery, F.E. and Trist, E.L. (1960), Sociotechnical systems. In C.W. Churchman and M. Verhulst (Eds.), *Management Science* (Vol. 2). Oxford, U.K.: Pergamon Press.
- Foster, M. (1972), An introduction to the theory and practice of action research in work organizations. *Human Relations*, 25, 529–556.
- Giddens, A. (1984), *The Constitution of Society*. Berkeley, CA: University of California Press.
- Greenbaum, J. and Kyng, M. (Eds.) (1991), *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, NJ: LEA Press.
- Griffin, T.G.C., Young, M.S. and Stanton, N.A. (2010), Investigating accident causation through information network modelling. *Ergonomics*, 53(2), 198–210.
- Heller, F. (1997), Sociotechnology and the environment. *Human Relations*, 50(5), 605–624.
- Hendrick, H.W. (1997), Organizational design and macroergonomics. In G. Salvendy (Ed.), *Handbook of Human Factors* (2nd Edn.). New York: John Wiley & Sons.
- Hendrick, H.W. and Kleiner, B.M. (2000), *Macroergonomics: An Introduction to Work System Design*. Santa Monica, CA: Human Factors and Ergonomics Society.
- Hollnagel, E. (March 2012), *A Birds Eye View of Resilience Engineering*. Loughborough, U.K.: Loughborough University.
- Hollnagel, E., Woods, D.D. and Leveson, N. (Eds.) (2006), *Resilience Engineering*. Farnham, U.K.: Ashgate.
- Hutchins, E. (1995), *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Karsh, B. (2003), Meso-ergonomics: A new paradigm for macroergonomics research. *Proceedings of the International Ergonomics Association 2006 Congress*, Maastricht, the Netherlands/Satan Barbara, CA: IEA Press.
- Karsh, B.-T. and Brown, R. (2010), Macroergonomics and patient safety: The impact of levels on theory, measurement, analysis, and intervention in patient safety research. *Applied Ergonomics*, 41(5), 674–681.

- Karsh, B.-T., Waterson, P.E. and Holden, R. (2014), A framework for mesoergonomic inquiry. *Applied Ergonomics*, 45(1), 45–54.
- Katz, D. and Kahn, R. (1966), *The Social Psychology of Organizations*. New York: John Wiley & Sons.
- Kelly, J. (1978), A reappraisal of sociotechnical systems theory. *Human Relations*, 31(12), 1069–1099.
- Kelly, J. (1982), *Scientific Management, Job Re-Design and Work Performance*. London, U.K.: Academic Press.
- King, S.D.M. (1964), *Training within the Organization*. London, U.K.: Tavistock Institute.
- Landauer, T.K. (1995), *The Trouble with Computers*. Cambridge, MA: MIT Press.
- Latour, B. (1992), Where are the missing masses? The sociology of a few mundane artefacts. In W.E. Bijker and J. Law (Eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press.
- Latour, B. (1999), *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
- Mayhew, D.J. (1999), *The Usability Engineering Lifecycle*. San Francisco, CA: Morgan Kaufmann.
- Maznevski, M.L. and Chudoba, K.M. (2000), Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science*, 11(5), 473–492.
- Medsker, G.J. and Campion, M.A. (1997), Job and team design. In G. Salvendy (Ed.), *Handbook of Human Factors* (2nd Edn.). New York: John Wiley & Sons.
- Moray, N. (2007), Real prediction of real performance. In J.R. Wilson, B. Norris, T. Clarke and A. Mills (Eds.), *People and Rail Systems: Human Factors at the Heart of the Railway*. London, U.K.: Ashgate.
- Mumford, E. (1995), *Effective Systems Design and Requirements Analysis: The ETHICS Method*. London, U.K.: Macmillan.
- Mumford, E. and Axtell, C.M. (2002), Tools and methods to support the design and implementation of new work systems. In D. Holman, T.D. Wall, C.W. Clegg, P. Sparrow and A. Howard (Eds.), *The New Workplace: A Guide to the Human Impact of Modern Working Practices*. London, U.K.: John Wiley & Sons.
- Mumford, E. and Hendricks, R. (1996), Business process re-engineering RIP. *People Management*, May 1996, 22–29.
- Nadin, S.J., Waterson, P.E. and Parker, S.K. (2001), Participation in job redesign: An evaluation of the use of a sociotechnical tool and its impact. *Human Factors and Ergonomics in Manufacturing*, 11(1), 53–69.
- Nielson, J.K. (1997), Usability testing. In G. Salvendy (Ed.), *Handbook of Human Factors* (2nd Edn.). New York: John Wiley & Sons.
- Older, M.T., Waterson, P.E. and Clegg, C.W. (1997), A critical assessment of task allocation methods and their applicability. *Ergonomics*, 40, 151–171.
- Parker, S.K. and Wall, T.D. (1998), *Job and Work Design: Organizing Work to Promote Well-Being and Effectiveness*. London, U.K.: Sage Publications.
- Preece, J., Rogers, Y. and Sharp, H. (2002), *Interaction Design: Beyond Human-Computer Interaction*. London, U.K.: John Wiley & Sons.
- Rasmussen, J. (1997), Risk management in a dynamic society: A modelling problem. *Safety Science*, 27, 183–213.
- Rasmussen, J. and Svedung, I. (2000), *Proactive Risk Management in a Dynamic Society*. Karlstad, Sweden: Swedish Rescue Services Agency.
- Rice, A.K. (1958), *Productivity and Social Organization*. London, U.K.: Tavistock Institute.
- Salmon, P.M., Stanton, N.A., Lenne, M., Jenkins, D.P., Rafferty, L. and Walker, G.H. (2011), *Human Factors Methods and Accident Analysis*. Farnham, U.K.: Ashgate.
- Symon, G.J. and Clegg, C.W. (1991a), Technology-led change: A study of the implementation of CAD/CAM. *Journal of Occupational Psychology*, 64, 272–290.
- Symon, G.J. and Clegg, C.W. (1991b), Implementation of a CAD/CAM system: The management of change at EML. In K. Legge, C.W. Clegg and N.J. Kemp (Eds.), *Case Studies in Information Technology, People and Organisations*. Oxford, U.K.: NCCC Blackwell.
- Trist, E.L. and Bamforth, K. (1951), Some social and psychological consequences of the long-wall method of coal-getting. *Human Relations*, 4, 3–38.
- Trist, E.L., Susman, G. and Brown, G.W. (1977), An experiment in autonomous group working in an American underground coal mine. *Human Relations*, 30, 201–236.
- van Eijnatten, F.M. (1993), *The Paradigm that Changed the Workplace*. Stockholm, Sweden: Swedish Centre for Working Life (Arbetslivcentrum).
- Walker, G.H., Stanton, N.A., Baber, C., Wells, L., Gibson, H., Salmon, P.M. and Jenkins, D.P. (2010), From ethnography to the EAST method: A tractable approach for representing distributed cognition in air traffic control. *Ergonomics*, 53(2), 184–197.

- Wall, T.D., Kemp, N.J., Jackson, P.R. and Clegg, C.W. (1986), An outcome evaluation of autonomous work groups: A long-term field experiment. *Academy of Management Journal*, 29, 280–304.
- Walton, R.E. (1977), Work innovations at Topeka: After six years. *Journal of Applied Behavioural Science*, 13, 422–433.
- Waterson, P.E. (2014), Health information technology and sociotechnical systems: A progress report on recent developments within the English NHS. *Applied Ergonomics*, 45(2), 150–161.
- Waterson, P.E., Clegg, C.W., Bolden, R.I., Pepper, K., Warr, P.B. and Wall, T.D. (1999), The use and effectiveness of modern manufacturing practices: A survey of U.K. industry. *International Journal of Production Research*, 37, 2271–2292.
- Waterson, P.E., Glenn, Y. and Eason, K. (2012), Preparing the ground for the ‘paperless hospital’: A case study of medical records management in a U.K. outpatient services department. *International Journal of Medical Informatics*, 81, 114–129.
- Waterson, P.E., Older Gray, M.T. and Clegg, C.W. (2002), A sociotechnical method for designing work systems. *Human Factors*, 44(3), 376–391.
- Weisbord, M.R. (2012), *Productive Workplaces: Dignity, Meaning and Community in the 21st Century*. New York: Jossey-Bass.
- Wilson, J.R. (2012), Fundamentals of systems ergonomics. *Work*, 41, 3861–3868.
- Wilson, J.R. and Haines, H.M. (1997), Participatory ergonomics. In G. Salvendy (Ed.), *Handbook of Human Factors* (2nd Edn.). New York: John Wiley & Sons.
- Wilson, J.R., Ryan, B., Schock, A., Ferreira, P., Smith, S. and Pitsopoulos, J. (2009), Understanding safety and production risks in rail engineering planning and production. *Ergonomics*, 52(7), 774–790.
- Zohar, D. and Luria, G. (2005), A multilevel model of safety climate: Cross-level relationships between organization and group-level climates. *Journal of Applied Psychology*, 90(4), 616–628.



---

# 29 Teamwork and Team Performance Measurement

*Alyssa Marshall, Tiffany Bisbey and Eduardo Salas*

## CONTENTS

|                                                  |     |
|--------------------------------------------------|-----|
| Team Performance Measurement.....                | 774 |
| What Is Teamwork? .....                          | 774 |
| Cooperation .....                                | 774 |
| Conflict.....                                    | 775 |
| Coordination.....                                | 775 |
| Communication .....                              | 775 |
| Coaching .....                                   | 776 |
| Cognition.....                                   | 776 |
| Composition .....                                | 776 |
| Context .....                                    | 777 |
| Culture.....                                     | 777 |
| Measuring Team Performance .....                 | 777 |
| Use Measures Based on Theory.....                | 778 |
| Measure Processes and Outcomes .....             | 778 |
| Measure Team Performance at Multiple Levels..... | 778 |
| What Should You Use to Measure Teams? .....      | 779 |
| Attitudes/Affect.....                            | 779 |
| Rating Scales .....                              | 779 |
| Behaviours.....                                  | 779 |
| Rating Scales .....                              | 780 |
| Checklists .....                                 | 780 |
| Frequency Scales.....                            | 780 |
| Event-Based Measurement Scales.....              | 780 |
| Automated Performance Measurement .....          | 781 |
| Cognitions .....                                 | 781 |
| Concept Mapping/Cognitive Networks .....         | 781 |
| Probed Protocol Analysis .....                   | 782 |
| Current Trends and Future Directions.....        | 782 |
| Human Performance Models.....                    | 782 |
| Simulation Use and Fidelity .....                | 783 |
| Dynamic Measurement Tools.....                   | 783 |
| New Framework Schemas.....                       | 783 |
| Relative Accuracy Index and Time Windows.....    | 783 |
| Psychometric Contributions .....                 | 784 |
| Conclusion .....                                 | 784 |
| References.....                                  | 784 |

## TEAM PERFORMANCE MEASUREMENT

Over the past several decades, there has been a shift in structure of many organisations. Workplaces that were once characterised by independent workers are now transitioning to systems of interdependent teams. As global competition increases, new pressures are affecting the way organisations do business. To successfully navigate the market, they need specific skills and expertise, as well as the ability to adapt quickly and effectively in uncertain situations (Kozlowski and Bell, 2003). This increasing complexity has led to team-based structures being used in many modern organisations (Salas et al., 2008a).

As the use of teams in the workplace became more common, the number of studies of teams increased (Kozlowski and Bell, 2003; Salas et al., 2008a; Van Hootegem et al., 2005). When investigators examined the use of teams within organisations, it became apparent that the use of teams did not automatically make them effective (Salas et al., 1995). Over the past 30 years, a wealth of research has emerged on how to compose, structure, manage, measure and promote team performance (Salas et al., 2008a).

When investigators decide to study team performance, there are a number of questions that must be answered (e.g. what constitutes good team performance? How will team performance be measured? What purpose will the team performance measurement serve?) The aim of this chapter is to address some of the concerns and issues associated with measuring team performance, as well as to provide a basic overview of the tools used to do so. We will define teamwork and team performance, outline relevant considerations for measuring team performance, detail team performance measurement tools and discuss the future of team measurement. This chapter aims to provide an update to older discussions of these topics (see Brannick et al., 1997; Salas et al., 2005a).

## WHAT IS TEAMWORK?

A team is defined as two or more individuals who have specific roles, work interdependently, are adaptable and share a common goal (Salas et al., 1992). Teams can vary widely on many factors, including their composition, organisation, objectives, lifespan, communication and task interdependence. Every member on a team has individual responsibilities that include both taskwork and teamwork.

Taskwork includes specific tasks designated to individual team members (Salas et al., 2014). These responsibilities can be completed individually and do not depend on any interaction with the rest of the team. In contrast, teamwork components depend on multiple team members communicating, coordinating and cooperating together (Salas et al., 2008a). Teamwork is essentially the set of interrelated knowledge, skills and attitudes (i.e. KSAs) that need to keep the team functioning as a unit (Baker, Day and Salas, 2006). Both taskwork and teamwork are necessary, as deficits in either will be detrimental to team performance (Salas et al., 2014).

A recent article by Salas et al. (2014) defined several critical teamwork considerations that impact team outcomes. Six of these considerations are core processes and emergent states of teamwork: (1) cooperation, (2) conflict, (3) coordination, (4) communication, (5) coaching and (6) cognition. All six of these factors are interdependent, interrelated and essential for successful team performance (Salas et al., 2013). In addition, three considerations are influencing conditions that impact the ways in which teams engage in teamwork: (7) composition, (8) culture and (9) context. Together, these factors can be known as the 9 Cs of teamwork.

## COOPERATION

Cooperation is the team competency that includes helping others, pacing work to meet team needs and behaving in a way that avoids misinterpretations within the team (Salas et al., 2004). There are a number of attitudes, beliefs and feelings that influence cooperation within teams and lead to better team outcomes. These include collective efficacy and trust within teams (Salas et al., 2014).

Collective efficacy, a team's shared perception of their ability to successfully perform tasks and meet goals (Bandura, 1997), is related to overall team performance and individual team member performance (Gully et al., 2002). Similarly, trust, which is based on the perceived intentions of others and expectations for their behaviour, is associated with perceived task performance and team satisfaction (Costa et al., 2001).

## CONFLICT

Conflict is the awareness of incompatible interests, beliefs or views (Jehn, 1997). Jehn and colleagues (Jehn, 1995; Jehn and Bendersky, 2005; Jehn, Greer, Levine and Szulanski, 2008) outline three separate types of conflict that can disrupt teams. *Relationship conflicts* are personal and not directly related to the task. These might include personality differences, political views, workplace gossip and hobbies. *Task conflicts* are related to differences in opinions regarding team tasks and individual subtasks. A third type, *process conflict*, relates to how to delegate tasks, resources and responsibilities within the team.

There is a lack of consensus regarding the effects of conflict on team performance, particularly concerning task conflict. Meta-analyses have found differing results, some supporting the notion that all types of conflict are negatively related to team performance (De Dreu and Weingart, 2003), while others suggest that task conflict has little or no effect on team performance (de Wit et al., 2012). Further, others have found that task conflict is beneficial, but only under certain conditions (Bradley et al., 2013), or that task conflict promotes the discussion of different viewpoints regarding the completion of the task, and that this discussion ultimately leads to better team performance (Jehn, 1997).

## COORDINATION

Coordination is the integration of individual operations in a way that achieves timely and satisfactory team performance (Eccles and Tenenbaum, 2004). It includes both the cognitive and behavioural mechanisms needed to perform a task, without which successful team performance would not be possible (Sims and Salas, 2007). Coordination can be either explicit, with team members intentionally utilising tools to integrate their operations (e.g. communicating and planning), or implicit, with team members anticipating other's needs and actions and adjusting their behaviour accordingly (Rico et al., 2008). A meta-analysis of studies examining team design features and performance in natural settings revealed that coordination within a team is positively related to team performance, with a stronger relationship present for teams engaged in knowledge-related tasks than physical tasks (Stewart, 2006).

## COMMUNICATION

Broadly, communication can be considered as all ways in which one mind affects another, whether verbal or non-verbal, intentional or unintentional (Eccles and Tenenbaum, 2004; Shannon and Weaver, 1949). However, such a definition may downplay the intricacies of the communication process. When studying team communication, it is important to consider the 'internal and external factors that may influence information sending, interpretation, and response' (Salas et al., 2014). Effective communication is highly related to other team processes as they would not be possible without information sharing and transfer between team members. Communication facilitates reaching and maintaining coordination (Ellingsen and Ostling, 2010; Miller and Moser, 2003), and in some situations, it may help resolve team conflict (Bornstein and Gilula, 2003). Because of interdependency between members, none of the other core processes and emergent states could exist without intra-team communication, nor could teamwork itself.

The impact communication has on team performance has been widely studied (Salas et al., 2014). There is evidence for a curvilinear relationship between communication and performance,

indicating that either too little or too much discussion may hurt performance (Patrashkavo-Bolzdoska et al., 2003). More important than the fact that teams are simply communicating may be factors like the relevance and accuracy of the message to the team task. A recent meta-analysis found that information sharing positively predicted team performance, particularly when unique information was shared (Mesmer-Magnus and DeChurch, 2009).

Research in the medical industry has highlighted the sometimes dangerous effects of poor communication. Observational studies reveal that disruptions in the team communication process are the strongest predictor of surgical errors, surpassing disruptions due to lack of resources, equipment issues, training of junior staff members and external interruption (Wiegmann et al., 2007). In fact, Lingard et al. (2004) found that about 30% of all procedurally relevant communications between medical team members contain an error. Of these errors, about a third resulted in visible effects on operating room conditions, including inefficiency, team tension, resource waste, workaround, delay, patient inconvenience and procedural error.

## COACHING

Coaching can be defined as interaction with a team intended to improve team processes and ultimately team performance (Wageman, 2001). Hackman and Wageman (2005) define three basic coaching functions: to increase team effort and commitment, improve team performance strategy and increase team knowledge and skill levels. A number of different coaching behaviours may fall within these bounds, including providing training and instruction, giving feedback, providing social support and making decisions for the team. Although these coaching behaviours are usually the responsibility of a team leader, they are certainly not the only responsibilities of a team leader. Leaders are also tasked with a variety of non-coaching behaviours, such as structuring teams, obtaining and organising team resources and facilitating work completion (Hackman and Wageman, 2005).

A recent study of 137 work teams in Taiwan has confirmed the importance of coaching in the workplace. Researchers found that team coaching is positively related to team processes, which in turn increase team effectiveness (Liu et al., 2009). In comparison, a separate study found that when team leaders and members were asked to rank the amount of the attention given to leadership activities, team coaching came in last (as cited in Hackman and Wageman, 2005). Taken together, these studies highlight a troubling discrepancy between research and practice in this area.

## COGNITION

The study of team cognition seeks to understand the relationship between the internal cognitive processes of individual team members and the external social processes they engage in as part of a team (Rosen et al., 2009). Simply, team cognition can be defined as the shared team knowledge and understanding that allows a team to successfully perform their task (Salas et al., 2014). This shared knowledge includes both static and dynamic knowledge and can be further broken down into task-related, team-related, process-related and goal-related knowledge (Wildman et al., 2012). When shared knowledge is ill-developed, it can potentially lead to teamwork failures and negative team outcomes (Salas et al., 2014). Further, meta-analytic results indicate that team cognition is positively related to team behavioural processes, motivational states and team performance (DeChurch and Mesmer-Magnus, 2010).

## COMPOSITION

It is important to consider the individuals that comprise a team when studying team performance. A team's composition is the set of individual attributes present within a team (Bell, 2007; Levine and Moreland, 1990). This includes both surface-level variables which can be reasonably assumed

after brief exposure (e.g. race, age and gender) and deep-level variables that require more knowledge of the individual (e.g. personality, values and attitudes; Bell, 2007).

Research has shown that diversity of task-related characteristics (e.g. expertise, education and organisation tenure) is positively related to both the quantity and the quality of team performance, while diversity of bio-demographic characteristics (e.g. gender, race and ethnicity) has no effect (Horwitz and Horwitz, 2007). This may be because as teams collaborate over time, the effects of demographic diversity weaken, but the effects of psychological diversity increase (Harrison et al., 2002). Beyond diversity, field studies show that good team performance is generally related to certain attributes, such as agreeableness, conscientiousness, openness to experience, collectivism and preference for teamwork (Bell, 2007).

## CONTEXT

The influence context has on behaviour and performance is widely recognised (Matsumoto, 2007; Patterson et al., 2004). Organisational context can be defined as the situational factors that affect behaviour in the workplace (Salas et al., 2014). Johns (2006) separates context into two levels of analysis, the omnibus and discrete contexts. The omnibus context includes the many defining characteristics of a situation, like the people, location, time and rational. Within the omnibus context is the discrete context, which includes the specific situational variables that influence behaviour. These situational variables can be associated with the task (e.g. autonomy and uncertainty), the social environment (e.g. social structure and social influence) or the physical environment (e.g. temperature and light; Johns, 2006).

Of particular interest are contextual variables that might negatively impact team performance. This includes organisational climate, the perception of an organisation's policies and practices (Patterson et al., 2004). Research has indicated that organisational climate affects individual performance (Parker et al., 2003) and the intention to share knowledge within a team (Bock et al., 2005).

Perhaps even more influential than organisational climate is a stressful environment. Stress occurs when an individual perceives a task or situation to be too demanding to be completed with their limited resources (Salas et al., 1996). When teams have to perform under acute stress (i.e. sudden, novel and intense stress), team performance suffers (Ellis, 2006). This is concerning for many teams, including those in medical, military, aviation and emergency response fields, where failures to perform can have potentially deadly consequences.

## CULTURE

Culture can be described as 'a blueprint to show individuals how to perceive, think, and act in a social environment' (Salas et al., 2008, p. 115). It is common to those of a specific people group (e.g. team, organisation, nation and region) and manifests itself in the values, norms and behaviours of individuals (Salas et al., 2014). Culture has become particularly relevant, as increasing globalisation in business has necessitated the use of multicultural work teams (Rockstuhl and Ng, 2008).

Brett et al. (2006) identify four areas that can derail a multicultural team's success: differing communication styles, differing attitudes towards authority, different approaches to decision making and language barriers. Meta-analytic findings suggest that cultural diversity has both positive and negative effects, leading to decreased social integration and increased task conflict, but also increased creativity and satisfaction (Stahl et al., 2010).

## MEASURING TEAM PERFORMANCE

At the core of all teamwork and team performance research are the methods through which team performance is evaluated. Team performance measurement provides a way to systematically examine the factors that cause effective or ineffective team performance (Rosen et al., 2008b).

Although they are essential, good measures of performance are not easy to develop (Umble and Umble, 2012). Several recent studies by Salas, Rosen and colleagues (Rosen et al., 2008b; Salas et al., 2009a,b) have defined best practices for measurement within simulation-based team training. Later we adapt a few of their recommendations beyond the realm of training to serve researchers looking to measure team performance for all purposes.

### USE MEASURES BASED ON THEORY

Research in the last several decades has produced a wealth of team-related knowledge (Salas et al., 2008a). When measuring team performance, this knowledge should be used to help focus attention on the aspects that matter most in achieving effective performance (Rosen et al., 2008b). It is important to consider all the KSAs that may affect your team's ability to complete their tasks (Salas et al., 2007b). This means examining both taskwork and teamwork demands.

Of course, the necessary taskwork-related KSAs will be highly variable depending on the team task at hand and the individual member's role. However, when it comes to teamwork-related KSAs, at least some (e.g. those related to communication and coordination) appear to be relevant to all types of teams (Salas et al., 2005b, 2007b). For this reason, frameworks describing general teamwork features can provide a good overview of the factors that should be considered for measurement (Rosen et al., 2008b). The *9 Cs of Teamwork* is a recent example of a general teamwork framework, although several other frameworks exist (see Cannon-Bowers et al., 1995; Dickinson and McIntyre, 1997; Marks et al., 2001; McIntyre and Salas, 1995; Salas et al., 2005b).

In addition to general models that apply to all types of teams, there may be some field-specific theory available in areas where teamwork research has been plentiful (e.g. health care and aviation). Frameworks tailored to an individual field can provide further insight into team performance measurement in that area (Rosen et al., 2008b). Rosen et al. (2008b) recommend considering a field-specific theory when available. Further, they emphasise the importance of using an empirically based theory regardless of whether or not it is domain specific.

### MEASURE PROCESSES AND OUTCOMES

When designing measures, it is important to clearly define what is being measured (Salas et al., 2009b). In the case of team performance, this should include the behaviours that occur both during a team performance episode and the outcomes that occur as a result of those behaviours (Rosen et al., 2008b). When only processes are measured, it is impossible to link them to positive or negative outcomes. Alternatively, when only outcomes are measured, it is impossible to know what behaviours led to the team's good or bad performance (Salas et al., 2009a). Together, process and outcome measures can be used to investigate the causes of poor team performance, and once causes are established, to inform corrective action (Rosen et al., 2008b).

### MEASURE TEAM PERFORMANCE AT MULTIPLE LEVELS

Team performance should be measured at both the individual and the team level (Rosen et al., 2008b; Salas et al., 2009a). Because of the interdependency associated with teamwork, team performance is ultimately determined by both individual and team-level factors (Rosen et al., 2008b). If all measures are taken at the team level, it is impossible to detect deficiencies in particular team members' performance. Conversely, if all measures are taken at the individual level, it is impossible to detect teamwork deficiencies (Rosen et al., 2008b). When performance is measured on both levels, the source of poor performance can be traced and corrective action can be taken (Rosen et al., 2008b).



## WHAT SHOULD YOU USE TO MEASURE TEAMS?

In the interest of improving team effectiveness, researchers use many different tools to measure team processes and outcomes. The competencies and constructs related to team performance can be categorised by the ABCs of teamwork: attitudes, behaviours and cognitions (Salas et al., 2008b). It is important to distinguish which category a construct belongs to when attempting to design a valid and reliable measurement tool. The multitude of different tools that have been created to assess team performance can be overwhelming, so we divide the tools into these categories based on their primary assessment target, though many can be used to measure a combination of competencies. As we noted, the best tools measure individual- and team-level variables, providing both outcomes and underlying team processes (DeChurch and Haas, 2008; DeShon et al., 2004; LePine et al., 2008; Rosen et al., 2011).

### ATTITUDES/AFFECT

Attitudes, or affective factors, refer to what team members feel and believe (Morgan et al., 1994). Attitudes are important to team performance researchers because they reflect an individual's internal state, which affects team interactions. Commonly measured attitudes in team assessment are collective orientation, collective efficacy, psychological safety, satisfaction, cohesion, trust, empowerment and goal commitment (Salas et al., 2008a). Team member attitudes use self-reported tools, usually utilising a simple Likert scale (see Chapter 4).

### Rating Scales

Rating scales typically run on successive continuums with equal intervals, each denoting an indicated level of a specific affective competency. Different variations of rating scales include the addition of descriptive words, numbers and/or a brief description (Robbins and Judge, 2008; also, see Chapter 4 of this book). To develop an attitudinal rating scale, the first step is to define the variable of interest (e.g. cohesion). The next step is to use the definition to develop statements or questions that will assess the construct. These statements are what participants will use to provide their personal ratings or attitudes. After developing these statements, a subject matter expert (SME) should be consulted to determine whether the statements adequately reflect the attitudinal variable in question. After adjusting the scales based on SME feedback, the final step is to administer the scale (Shultz and Whitney, 2005).

Rating scales are the most common method of team performance measurement due to the ease of data collection and the low cost. Measures of attitude can provide significant contributions to team assessment, though results should be examined with caution. Attitudes reflect how an individual feels, therefore are measured subjectively and more susceptible to biases (Smith-Jentsch et al., 2013).

### BEHAVIOURS

Behaviours refer to the actions of a team member that affect team performance (Morgan et al., 1994). Commonly measured behavioural competencies include monitoring, adapting, backup behaviour, coordination, leadership, communication, mission analysis, planning, conflict resolution and intra-team feedback (Salas et al., 2008a). As opposed to the self-reported affect measures, the source of behavioural information can be a variety of outside observers, such as supervisors, other team members, SMEs or automated systems, in addition to self-report (Smith-Jentsch et al., 2013). As there are more sources of behavioural performance data, there are also more methods available used to collect these data.

## Rating Scales

Behavioural rating scales follow the same concept and development procedure as the attitudinal rating scales discussed in the previous section. These scales allow for a participant to rate team skills and behavioural competencies that are indicative of team performance. The BARS (Behaviourally Anchored Rating Scales) method of team behaviour assessment, created by Smith and Kendall (1963), comprises rating scales that utilise behavioural anchors alongside a numbered scale to denote a team member's skill level. The anchors usually consist of a descriptive word (e.g. poor, average, excellent) along with a definition and examples or criteria that constitute each respective level (Wildman et al., 2011). These anchors are fixed to control for any individual differences between coders, as team member behaviours must meet the specified criteria, and therefore are considered more reliable than other types of rating scales (Salas et al., 2008b).

Rating behavioural performance is also referred to as behavioural coding. To measure behavioural data, a coder is required to observe participant behaviour. The coder must be trained to analyse behaviours of interest using the measurement tool. This is essential to all behavioural coding because the observer is essentially creating the experimental data. In order for the data to be valid, the tool must be interpreted how it was originally intended and not at the discretion of the coder (Salas et al., 2008b). It is also important to note that behavioural rating scales are most appropriate for use when performance quality is more imperative than performance quantity (Smith-Jentsch et al., 2013).

## Checklists

A simple tool to measure team performance is a checklist. Checklists are reserved for tasks that have an explicit procedure to follow. A typical checklist has items that must be completed for team success, and a rater must simply mark if the behaviour occurs or not. Checklists are most appropriate when failures to complete necessary tasks are more important to team success than completing unnecessary tasks (Smith-Jentsch et al., 2013).

## Frequency Scales

Another simple measurement of team performance is a frequency count. Like checklists, frequency scales require the rater to observe a team for behaviours of interest. This tool varies in that the observer must tally all occurrences. Accordingly, this measure is most appropriate when quantity characterises team effectiveness rather than quality (Smith-Jentsch et al., 2013). For example, a frequency scale would be useful in an operating room simulation to assess how often a surgeon's decisions are challenged; and a checklist would be useful to assess if the proper steps in the procedure were followed by a surgeon.

## Event-Based Measurement Scales

Event-based measurement scales are much more elaborate than other rating scales. These scales consist of embedding latent triggers, intended to elicit participant behaviours, into team scenarios while being observed for behavioural response. The observing raters often have a list of cues or qualifying behaviours to look for, then record whether or not the targeted behaviour was performed. This means that the observers do not have to be SMEs because the targeted behaviours have been deemed appropriate by SMEs in the design. Because raters will know what triggers to look for and when they will occur, it reduces potential ambiguities in behavioural coding, which leads to increased reliability. Knowing exactly when the behaviour should be performed allows researchers to directly link measurement and the competency of interest (Rosen et al., 2008b). After embedding a specific trigger into a scenario, a script is then drafted in order to standardise the experimental method and maintain consistency through all participants and trials. Event-based tools are also useful in studying psychometric properties. They are frequently used to train and evaluate military teams due to the characteristic, regimented events in their line of work and predictable behavioural responses.

Examples of event-based measurement tools include SPOTLITE, SMARTER and TARGETS (see Fowlkes et al., 1994; Macmillan et al., 2013; and Rosen et al., 2008a for detailed information on these measurement tools). These tools are especially useful in measuring real-time team performance in simulation-based training environments that provide a variety of different methods to achieve a single successful outcome (Macmillan et al., 2013).

### Automated Performance Measurement

Automated performance measurement tools encompass a wide variation of systems. This includes anything from a spaceflight simulation to a motion sensor. These complex technical systems become increasingly more prevalent as technology advances, though they cannot yet fully replace non-automated measures; they should be used in conjunction with non-automated measures to increase reliability. Automated measures can be less obtrusive, less biased and administered more consistently than typical human non-automated measures of team performance, though they can only detect overt behavioural responses (Cooke et al., 2004).

### COGNITIONS

Cognitions refer to what team members think, know and perceive (Salas et al., 2007). Cognitive abilities are arguably the most important measures because they have the strongest and most consistently positive relationships with team performance (Bell, 2007; Devine and Phillips, 2001; Stewart, 2006). Similar to attitude measures, cognitions are often self-reported (Wildman et al., 2013). Further, the concepts of team attitudes and team cognitions are often confused. It will be helpful to remember that attitudes reflect *feelings* (e.g. 'I enjoy completing this task') and cognitions reflect *perceptions* (e.g. 'I think this task is easy'). Shared perceptions are vital to team success, meaning that ineffective team cognitions cause team failure, as tasks become more reliant on the performance of other team members (Mathieu et al., 2008). Because cognitions refer to abstract constructs, the tools used to measure them are more complex.

### Concept Mapping/Cognitive Networks

Concept maps were first introduced by Novak and Gowin (1984) as graphical representations of information networks that demonstrate the relationships between concepts in a specific area. Rather than a performance assessment tool, concept mapping is considered to be a knowledge elicitation tool (Cooke et al., 1997; Mohammed et al., 2000). A guideline to follow in measuring development is that an efficient knowledge elicitation tool should measure knowledge accuracy while also measuring the structure, content and essence of teams and their members (Cooke et al., 2000). Data come directly from team members through interviews and surveys, or indirectly through post hoc analyses. The process requires a participant to sort, place and link related concepts into a model, creating a visual representation of concept relationships (Burtscher et al., 2011). It essentially develops a model (or *cognitive network*) with components linked together based on their relationships with other components in the map. Concepts can be linked together for any reason from similarity to causality; for this reason, there are many different types of concept maps (Mohammed et al., 2000). Concept maps are often used to brainstorm or to organise thoughts in a graphical way, but are most often used in team performance measurement to assess mental models (Ellis, 2006; Marks et al., 2000, 2002; Wildman et al., 2013).

In addition to the different linkage types, there are also several technique variations in concept mapping. The first method involves the team member placing provided concepts into a provided model (Marks et al., 2002). In another method, the model is provided, but the participant fills it in with their own concepts. In the final method, the participant creates his or her own model, concepts and linkages (Wildman et al., 2013). All methods of concept mapping are useful to team performance research because the cognitive networks that they create show a great deal about the shared strategic thinking of teams, revealing the underlying knowledge network that leads to team success (Ellis, 2006; Marks et al., 2000, 2002).

Cognitive networks can also be developed using pattern analysis. A popular tool used is the Pathfinder scaling technique (Wildman et al., 2013). Pathfinder is a mathematical technique that elicits a team member's perception of similarities between concepts and produces linkages through representative scaling (Mohammed et al., 2000). Participants are presented with every possible combination of relevant concepts, one pair at a time, until they have rated each pair. Pathfinder then uses an algorithm that aggregates the provided ratings to construct a structural map that illustrates the linked constructs. Further, Pathfinder identifies the concepts with a high degree of similarity as direct links and those with a low degree of similarity as indirect links (Cooke et al., 2003). Pathfinder analysis only allows for the assessment of perceived similarities, and not causality, but paints a precise representation of shared knowledge with the benefit of linkage strength distinction (Wildman et al., 2013).

### **Probed Protocol Analysis**

In-task probes use simulations to demonstrate the shared knowledge of a team. In these measures, team members participate in a scenario that incorporates the competencies of interest, while being answering questions regarding perception (Wildman et al., 2013). In a typical probed protocol experiment, participants are given a set of steps necessary to complete a task and then asked to explain their choices regarding task selectivity, task sequence and other team members' concurrent responsibilities at various stopping points within the simulation (Bolstad and Endsley, 2003). This tool allows researchers to compare a team member's thought process with that of the other team members to get a better understanding of individual and team comprehension. The problem with this method is that it is highly obtrusive due to the frequent interruption during the simulation. A good way to mitigate this concern is to incorporate a confederate into the team that will periodically request perceptual information from other team members (Camelo-Ordaz et al., 2005). The main advantage of probed protocol analysis is that it collects information at multiple data points in real time, which allows the researcher to see a progression of cognition over the course of a scenario (Wildman et al., 2013).

## **CURRENT TRENDS AND FUTURE DIRECTIONS**

Team performance measures have expanded and adapted extensively in the last decade to accommodate for new and more prevalent types of teams and the interaction with automated systems. This includes the influx of globally dispersed virtual teams and the use of sociotechnical systems in daily operations and evaluations (Bullock and Klein, 2011). Technological advances change the nature of teams and their environments, making the assessment of team performance more dynamic and ambiguous. The technological increase has also led to fewer humans being used to complete tasks in areas where activities might be dangerous or beyond human physical or cognitive capability (e.g. the military and space exploration).

### **HUMAN PERFORMANCE MODELS**

Researchers are currently using human performance modelling to counteract threats to team effectiveness. Ergonomics and human factors (E/HF) research in this area is not new in itself; this application uses the concepts to develop systems that predict human behaviour based on models of mental activities, usually while humans are interacting with a system. In the team performance context, human performance models seek to analyse performance by depicting the major processes of a scenario and their relationship. These tools can be used to identify and evaluate the effectiveness of systems to predict errors (Foyle et al., 2005). Effective models learn as humans interact with them and gradually predict outcomes based on the processes of interest. Human performance models can predict how teams will perform in certain situations and how their different processes will interact when presented with certain cues. Models can then be used as experimental tools to recognise patterns in human cause-and-effect relationships, prompting future research opportunities that transfer to real-world team performance (Pharmer and Campbell, 2000).

### **Simulation Use and Fidelity**

Human performance modelling used in combination with simulations can be very useful in team training and the measurement of team performance. This trend of simulation use in team training and daily operations has dominated teams in recent years, especially those in health care, the military and other high-reliability organisations (Fowlkes et al., 1994; Macmillan et al., 2013; Rosen et al., 2008a). This progression towards simulated work environments and increased incorporation of technology requires adapted tools to measure team performance. Human performance models guide simulation design to enable the measurement of team performance in scenarios with high-fidelity (Pharmer and Campbell, 2000). Human performance models and their resulting simulations are being used more and more to create team and system designs, as well as to evaluate performance. Research is also often done to evaluate team member perceptions of simulation reality (Pandzic et al., 2001). Areas that currently utilise simulations and plan to continue capitalising on advanced simulated models in the future include space exploration missions, mission strategy, efficient health care operations, intelligent systems emulating human processes and optimal resource utilisation (see Chapter 8 for further discussion of simulation use within E/HF research).

### **Dynamic Measurement Tools**

Dynamic assessment is also an important issue in the assessment of team performance. For example, online training and assessment of teams can reduce the cost and travel time for operations as these systems are capable of being accessed worldwide (Solomon, 2001). Dynamic assessments refer to the real-time measurement of teams as they make decisions and perform moment to moment, a difficult task to assess due to the difficulties of isolating the process of interest to capture in real time. Further, data collection of some teams is often limited due to restricted access of remote measurement for safety and security reasons (e.g. government and military teams), as well as restrictions of video or audio recording team interactions (Kaber et al., 2006). Such recording could become more feasible in the near future with the recent advancements in smartphone technologies and reduced sensor size. These developments come with less obtrusive measurements and reduce the need for face-to-face observation.

### **New Framework Schemas**

Other researchers have developed a model for diagnosing and solving performance problems of teams in a total quality system that provides feedback for continuous improvement of processes (Çiçek et al., 2005). This model is unique because it methodically considers the input quality of other teams and external influences in evaluations. It also interactively involves the team in the measurement of its own performance and feedback. This model provides a way for teams to recognise and act on current problems by analysing team structure, inputs, processes and outputs of performance separately and continuously. It is particularly useful to senior management as it is able to oversee the status of all interdependent teams and create training programmes accordingly (Çiçek et al., 2005). In many organisations, a team's effectiveness is largely influenced by the input quality to team processes (Bell, 2007). Its effective usage requires an organisational, collectivistic culture that stresses cooperation among teams towards the organisation's goals and mission. Though this tool has been found to be effective and rated satisfactorily in health-care teams, it has not been fully tested for all forms of validity and reliability; therefore, more research should be conducted before applying this method in other experimental settings (Çiçek et al., 2005).

### **Relative Accuracy Index and Time Windows**

Recently, researchers have developed a tool for the evaluation of team performance using the relative accuracy index (RAI), a temporal accuracy measure, to determine appropriate ratings based on individual task demands (Thiruvengada and Rothrock, 2007). This approach was developed as a way to avoid biases between raters, as well as between SMEs in determining rater procedures. It addresses challenges in team performance measurement by creating evaluation guidelines for

teams based on time windows and RAI to more accurately map processes with outcomes. This tool was tested in a simulated military air traffic control environment and therefore should only be generalised to the *command and control* context (Thiruvengada and Rothrock, 2007). Additionally, it has only been tested to assess backup behaviour, making further testing necessary for practical generalisability.

### Psychometric Contributions

Discovering the predictive utility of rating scales greatly adds to psychometric research, as well as to the designing of tools used to measure team performance. Kline (2003) has begun studying the psychometric properties of scales currently used to predict performance. She analysed responses from scales measuring team leadership behaviours and mission orientation with perceptions of team performance to find that perceived performance is related equally to leadership and orientation and that leadership behaviours are more correlated than orientation with team performance and satisfaction (Kline, 2003). Tests such as this validate dispositions in team performance research that recognise the influence of multilevel and context variables on performance outcomes and how these relationships may fluctuate with differing affective reactions of team members.

## CONCLUSION

The use of teams in the workplace has progressed significantly over the past few decades to encompass all facets of business, industry, health care and the military. The need for team performance measures is vastly evident in today's workforce and is maintained by industrial and organisational psychologists and human factors and ergonomics researchers. Though we have uncovered a great deal about the tools used to effectively measure team performance, there is a need to improve the translation of methods and evaluations to those who actually need them, mainly practitioners that would benefit from their applications. Practitioners may benefit from knowing how these tools were developed and exactly what they are designed to measure.

The tools discussed in this chapter, like all measures of team performance, must be applied across the lifespan of a team in order to be effective. Teams are constantly evolving and changing, therefore should be re-evaluated accordingly. While evaluating teams, it is absolutely necessary to assess the performance of individuals as well as the team as a whole. Most of all, these tools must be used diligently to ensure that all considerations and challenges are addressed, including outcome and process measures.

This chapter should be utilised as a guide to provide the reader with the foundations of teamwork and team performance measurements, as well as the complexities that exist within this facet of team research. Failures in critical team processes have devastating effects on team performance (Salas et al., 2008a). We hope to mitigate these risks and increase awareness within the fields of industrial and organisational psychology, human factors and ergonomics research and the students learning the vital concepts within these fields. By educating readers regarding these tools, their applications and the science behind their uses, it encourages the accurate usage of effective team performance measurement.

## REFERENCES

- Baker, D. P., Day, R. and Salas, E. (2006). Teamwork as an essential component of high reliability organizations. *Health Services Research*, 41(4), 1576–1598.
- Bandura, A. (1997). Collective efficacy. In A. Bandura (Ed.), *Self-Efficacy: The Exercise of Control* (pp. 477–525). New York: W.H. Freeman and Company.
- Bell, S. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *The Journal of Applied Psychology*, 92(3), 595–615.



- Bock, G. W., Zmud, R. W., Kim, Y. G. and Lee, J. N. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Quarterly*, 29, 87–111.
- Bolstad, C. A. and Endsley, M. R. (October 2003). Measuring shared and team situation awareness in the army's future objective force. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 3, pp. 369–373). Santa Monica, CA: HFES, SAGE Publications.
- Bornstein, G. and Gilula, Z. (2003). Between-group communication and conflict resolution in assurance and chicken games. *Journal of Conflict Resolution*, 47(3), 326–339.
- Bradley, B. H., Klotz, A. C., Postlethwaite, B. E. and Brown, K. G. (2013). Ready to rumble: How team personality composition and task conflict interact to improve performance. *Journal of Applied Psychology*, 98(2), 385.
- Brannick, M. T., Salas, E. and Prince, C. W. (Eds.) (1997). *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Brett, J., Behfar, K. and Kern, M. C. (2006). Managing multicultural teams. *Harvard Business Review*, 84(11), 84.
- Bullock, C. and Klein, J. T. (2011). Virtual work environments in the post-recession era. Irvine, CA: Brandman University.
- Burtscher, M. J., Kolbe, M., Wacker, J. and Manser, T. (2011). Interactions of team mental models and monitoring behaviours predict team performance in simulated anesthesia inductions. *Journal of Experimental Psychology: Applied*, 17(3), 257–269.
- Camelo-Ordaz, C., Hernández-Lara, A. B. and Valle-Cabrera, R. (2005). The relationship between top management teams and innovative capacity in companies. *Journal of Management Development*, 24, 683–705.
- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E. and Volpe, C. E. (1995). Defining competencies and establishing team training requirements. *Team Effectiveness and Decision Making in Organizations*, 333, 380.
- Çiçek, M., Köksal, G. and Özdemirel, N. (2005). A team performance measurement model for continuous improvement. *Total Quality Management & Business Excellence*, 16(3), 331–349.
- Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C. and Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research, and Practice*, 7, 179–199.
- Cooke, N. J., Salas, E., Cannon-Bowers, J. A. and Stout, R. (2000). Measuring team knowledge. *Human Factors*, 42, 151–173.
- Cooke, N. J., Salas, E., Kiekel, P. A. and Bell, B. (2004). Advances in measuring team cognition. In E. Salas and S. M. Fiore (Eds.), *Team Cognition: Understanding the Factors That Drive Process and Performance* (pp. 83–106). Washington, DC: American Psychological Association.
- Cooke, N. J., Stout, R. and Salas, E. (1997). Expanding the measurement of situation awareness through cognitive engineering methods. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 215–219). Santa Monica, CA: Human Factors and Ergonomics Society.
- Costa, A., Roe, R. A. and Taillieu, T. (2001). Trust within teams: The relation with performance effectiveness. *European Journal of Work and Organizational Psychology*, 10(3), 225–244. doi:10.1080/13594320143000654.
- DeChurch, L. A. and Haas, C. D. (2008). Examining team planning through an episodic lens: Effects of deliberate, contingency, and reactive planning on team effectiveness. *Small Group Research*, 39(5), 542–568.
- DeChurch, L. A. and Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology*, 95(1), 32.
- De Dreu, C. K. and Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology*, 88(4), 741.
- Devine, D. J. and Philips, J. L. (2001). Do smarter teams do better: A meta-analysis of cognitive ability and team performance. *Small Group Research*, 32(5), 507–532.
- DeShon, R., Kozlowski, S., Schmidt, A., Milner, K. and Weichmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *The Journal of Applied Psychology*, 89, 1035–1056.
- de Wit, F. R., Greer, L. L. and Jehn, K. A. (2012). The paradox of intragroup conflict: A meta-analysis. *Journal of Applied Psychology*, 97(2), 360.
- Dickinson, T. L. and McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. In M. T. Brannick, E. Salas, and C. Prince (Eds.), *Team Performance Assessment and Measurement: Theory Methods and Applications* (pp. 19–43). Mahwah, NJ: Lawrence Earlbaum Associates.
- Eccles, D. W. and Tenenbaum, G. (2004). Why an expert team is more than a team of experts: A social-conceptualization of team coordination and communication in sport. *Journal of Sport & Exercise Psychology*, 26(4), 542.

- Ellingsen, T. and Östling, R. (2010). When does communication improve coordination? *The American Economic Review*, 100(4), 1695–1724.
- Ellis, A. P. J. (2006). System breakdown: The role of mental models and transactive memory in the relationship between acute stress and team performance. *Academy of Management Journal*, 49(3), 576–589.
- Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T. and Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology*, 6, 47–61.
- Foyle, D. C., Hooley, B. L., Byrne, M. D., Corker, K. M., Deutsch, S., Lebiere, C. and Wickens, C. D. (September 2005). Human performance models of pilot behavior. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 12, pp. 1109–1113). Santa Monica, CA: HFES, SAGE Publications.
- Gully, S. M., Joshi, A., Incalcaterra, K. A. and Beaubien, J. (2002). A meta-analysis of team-efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships (abstract). *Journal of Applied Psychology*, 87(5), 819.
- Hackman, J. R. and Wageman, R. (2005). A theory of team coaching. *Academy of Management Review*, 30(2), 269–287.
- Harrison, D. A., Price, K. H., Gavin, J. H. and Florey, A. T. (2002). Time, teams, and task performance: Changing effects of surface-and deep-level diversity on group functioning. *Academy of Management Journal*, 45(5), 1029–1045.
- Horwitz, S. K. and Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 33(6), 987–1015.
- Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, 40(2), 256–282.
- Jehn, K. A. (1997). A qualitative analysis of conflict types and dimensions in organizational groups. *Administrative Science Quarterly*, 42(3), 530–557.
- Jehn, K. A. and Bendersky, C. (2003). Intragroup conflict in organizations: A contingency perspective on the conflict-outcome relationship. *Research in Organizational Behavior*, 25, 187–242.
- Jehn, K. A., Greer, L., Levine, S. and Szulanski, G. (2008). The effects of conflict types, dimensions, and emergent states on group outcomes. *Group Decision and Negotiation*, 17(6), 465–495.
- Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review*, 31(2), 386–408.
- Kaber, D. B., Riley, J. M., Hyatt, J. R. and Reynolds, J. P. (2006). Assessing infantry soldier situation awareness in virtual environment-based training of urban terrain operations. In *Proceedings of the International Ergonomics Association Triennial Conference* (pp. 1–6). Maastricht, the Netherlands: Elsevier.
- Kline, T. B. (2003). The psychometric properties of scales that assess market orientation and team leadership skills: A preliminary study. *International Journal of Testing*, 3(4), 321–332.
- Kozlowski, S. W. J. and Bell, B. S. (2003). Work groups and teams in organizations. In W. C. Borman, D. R. Ilgen and R. J. Klimoski (Eds.), *Handbook of Psychology: Industrial and Organizational Psychology* (Vol. 12, pp. 333–375). New York: Wiley.
- LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E. and Saul, J. R. (2008). A meta-analysis of team-work processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, 61(2), 273–307.
- Levine, J. M. and Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology*, 41(1), 585–634.
- Lingard, L., Espin, S., Whyte, S., Regehr, G., Baker, G. R., Reznick, R. and Grober, E. (2004). Communication failures in the operating room: An observational classification of recurrent types and effects. *Quality and Safety in Health Care*, 13(5), 330–334.
- Liu, C., Pirola-Merlo, A., Yang, C. and Huang, C. (2009). Disseminating the functions of team coaching regarding research and development team effectiveness: Evidence from high-tech industries in Taiwan. *Social Behavior and Personality*, 37(1), 41–58. doi:10.2224/sbp.2009.37.1.41.
- MacMillan, J., Entin, E. B., Morley, R. and Bennett Jr., W. (2013). Measuring team performance in complex and dynamic military environments: The SPOTLITE method. *Military Psychology*, 25(3), 266–279.
- Marks, M. A., Mathieu, J. E. and Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26(3), 356–376.
- Marks, M. A., Sabella, M. J., Burke, C. S. and Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology*, 87, 3–13.
- Marks, M. A., Zaccaro, S. J. and Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology*, 85, 971–986.

- Mathieu, J. E., Maynard, M. T., Rapp, T. and Gilson, L. L. (2008). Team effectiveness 1997–2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, 34(3), 410–476.
- Matsumoto, D. (2007). Culture, context, and behavior. *Journal of Personality*, 75(6), 1285–1320.
- McIntyre, R. M. and Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo, E. Salas and Associates (Eds.), *Team Effectiveness and Decision Making in Organizations* (pp. 9–45). San Francisco, CA: Jossey-Bass.
- Mesmer-Magnus, J. R. and DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, 94(2), 535.
- Miller, J. H. and Moser, S. (2003). Communication and coordination. *Complexity*, 9(5), 31–40.
- Mohammed, S., Klimoski, R. and Rentsch, J. R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods*, 3, 123–165.
- Morgan, B. B., Salas, E. and Glickman, A. S. (1994). An analysis of team evolution and maturation. *The Journal of General Psychology*, 120(3), 277–291.
- Novak, J. D. and Gowin, D. B. (1984). *Learning How to Learn*. Cambridge, U.K.: Cambridge University Press.
- Pandzic, I., Babski, C., Capin, T., Lee, W., Magnenat-Thalmann, N., Musse, S. R., Moccozet, L., Seo, H. and Thalmann, D. (2001). Simulating virtual humans in networked virtual environments. *Presence*, 10, 632–646.
- Parker, C. P., Baltes, B. B., Young, S. A., Huff, J. W., Altmann, R. A., Lacost, H. A. and Roberts, J. E. (2003). Relationships between psychological climate perceptions and work outcomes: a meta analytic review. *Journal of Organizational Behavior*, 24(4), 389–416.
- Patrashkova-Volzdoska, R. R., McComb, S. A., Green, S. G. and Compton, W. D. (2003). Examining a curvilinear relationship between communication frequency and team performance in cross-functional project teams. *IEEE Transactions on Engineering Management*, 50(3), 262–269.
- Patterson, M., Warr, P. and West, M. (2004). Organizational climate and company productivity: The role of employee affect and employee level. *Journal of Occupational and Organizational Psychology*, 77(2), 193–216.
- Pharmer, J. A. and Campbell, G. E. (2000). Verification and validation of three human performance models within the context of a field experiment. Retrieved May 15, 2002, from the Office of Naval Research International Field Office Website: <http://www.ehis.navy.mil/tb/humanscience/masakowski/dera/Nov00/SKI/Presentations/human%20perform%20modles%20pharmer%20&%20campbell.ppt>.
- Rico, R., Sánchez-Manzanares, M., Gil, F. and Gibson, C. (2008). Team implicit coordination processes: A team knowledge-based approach. *Academy of Management Review*, 33(1), 163–184.
- Robbins, S. P. and Judge, T. A. (2008). *Organizational Behavior* (13th edn., p. 752). Upper Saddle River, NJ: Prentice Hall.
- Rockstuhl, T. and Ng, K. Y. (2008). The effects of cultural intelligence on interpersonal trust in multicultural teams. In S. Ang and L. V. Dyne (Eds.), *Handbook of Cultural Intelligence: Theory, Measurement, and Applications* (pp. 206–220). Armonk, New York: M.E. Sharpe, Inc.
- Rosen, M. A., Bedwell, W. L., Wildman, J. L., Fritzsche, B. A., Salas, E. and Burke, C. S. (2011). Managing adaptive performance in teams: Guiding principles and behavioral markers for measurement. *Human Resource Management Review*, 21(2), 107–122.
- Rosen, M. A., Salas, E., Fiore, S. M., Pavlas, D. and Lum, H. C. (2009). Team cognition and external representations: A framework and propositions for supporting collaborative problem solving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 18, pp. 1295–1299). SAGE Publications.
- Rosen, M. A., Salas, E., Silvestri, S., Wu, T. S. and Lazzara, E. H. (2008a). A measurement tool for simulation-based training in emergency medicine: The simulation module for assessment of resident targeted event responses (SMARTER) approach. *Simulation Healthcare*, 3(3), 170–179.
- Rosen, M. A., Salas, E., Wilson, K. A., King, H. B., Salisbury, M., Augenstein, J. S. and Birnbach, D. J. (2008b). Measuring team performance in simulation-based training: Adopting best practices for health-care. *Simulation in Healthcare*, 3(1), 33–41.
- Salas, E., Bowers, C. A. and Cannon-Bowers, J. A. (1995). Military team research: 10 years of progress. *Military Psychology*, 7(2), 55.
- Salas, E., Cooke, N. J. and Rosen, M. A. (2008a). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50(3), 540–547.
- Salas, E., Dickinson, T. L., Converse, S. A. and Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey and E. Salas (Eds.), *Teams: Their Training and Performance* (pp. 3–29). Westport, CT: Ablex Publishing.

- Salas, E., Driskell, J. E. and Hughes, S. (1996). The study of stress and human performance. In J. E. Driskell and E. Salas (Eds.), *Stress and Human Performance* (pp. 1–45). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Salas, E., Lazzara, E. H., Benishek, L. E. and King, H. (2013). On being a team player: Evidence-based heuristic for teamwork in interprofessional education. *Medical Science Educator*, 23(3), 524–531.
- Salas E., Priest, H. A. and Burke, C. S. (2005a). Teamwork and team performance measurement. In J. R. Wilson and N. Corlett (Eds.), *Evaluation of Human Work* (pp. 793–808). Boca Raton, FL: Taylor & Francis Group.
- Salas, E., Rosen, M. A., Burke, C. S. and Goodwin, G. (2008b). The wisdom of collectives in organizations: An update of the teamwork competencies. In E. Salas, G. Goodwin and C. S. Burke (Eds.), *Team Effectiveness in Complex Organizations: Cross-Disciplinary Perspective and Approaches* (pp. 39–79). New York: Psychology Press.
- Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D. and Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine* (Special Supplement on Operational Applications of Cognitive Performance Enhancement Technologies), 78(5), B77–B85.
- Salas, E., Rosen, M. A., Held, J. D. and Weissmuller, J. J. (2009a). Performance measurement in simulation-based training: A review and best practices. *Simulation & Gaming*, 40(3), 328–376.
- Salas, E., Rosen, M. A., Weaver, S. J., Held, J. D. and Weissmuller, J. J. (2009b). Research on SBT leads to the development of guidelines applicable to diverse training scenarios. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 17(4), 12–18.
- Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L. and Lazzara, E. H. (2014). Understanding and improving teamwork in organizations: A scientifically based practical guide. Accepted for publication at *Human Resource Management*. doi: 10.002/hrm.21628.
- Salas, E., Sims, D. E. and Burke, C. S. (2005b). Is there a “big five” in teamwork? *Small Group Research*, 36(5), 555–599.
- Salas, E., Sims, D. E. and Klein, C. (2004). Cooperation at work. In C. D. Spielberger (Ed.), *Encyclopedia of Applied Psychology* (pp. 497–506). Boston, MA: Elsevier Academic Press.
- Salas, E., Stagl, K. C., Burke, C. S. and Goodwin, G. F. (2007). Fostering team effectiveness in organizations: Toward an integrative theoretical framework. In *Nebraska Symposium on Motivation* (Vol. 52, p. 185). Lincoln, NE: University of Nebraska Press.
- Salas, E., Wilson, K. A. and Lyons, R. (2008). Designing and delivering training for multicultural interactions in organizations. In D. L. Stone and E. F. Stone-Romero (Eds.), *The Influence of Culture on Human Resource Management Processes and Practices* (pp. 115–134). Hoboken, NJ: Taylor & Francis Group.
- Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, pp. 27–38.
- Shultz, K. S. and Whitney, D. J. (2005). *Measurement Theory in Action: Case Studies and Exercises*. Thousand Oaks, CA: SAGE Publications, Inc.
- Sims, D. E. and Salas, E. (2007). When teams fail in organizations: What creates teamwork breakdowns? In J. Langan-Fox, C. L. Cooper and R. J. Klimoski (Eds.), *Research Companion to the Dysfunctional Workplace: Management Challenges and Symptoms* (p. 302). Northampton, MA: Edward Elgar Publishing.
- Smith, P. C. and Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *The Journal of Applied Psychology*, 47, 149–155.
- Smith-Jentsch, K. A., Sierra, M. J. and Wiese, C. W. (2013). How, when, and why you should measure team performance. In E. L. Salas, S. I. Tannenbaum, D. Cohen and G. Latham (Eds.), *Developing and Enhancing Teamwork in Organizations* (pp. 552–580). Hoboken, NJ: Jossey-Bass.
- Solomon, C. M. (2001). Managing virtual teams. *Workforce*, 80(6), 60–65.
- Stahl, G. K., Maznevski, M. L., Voigt, A. and Jonsen, K. (2010). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies*, 41(4), 690–709.
- Stewart, G. L. (2006). A meta-analytic review of relationships between team design features and team performance. *The Journal of Management*, 32(1), 29–55.
- Thiruvengada, H. and Rothrock, L. (2007). Time windows-based team performance measures: A framework to measure team performance in dynamic environments. *Cognition, Technology & Work*, 9(2), 99–108.
- Umble, E. and Umble, M. (2012). Illustrating the impact of performance measurement systems on organizational performance: The blue-green game. *Decision Sciences Journal of Innovative Education*, 10(3), 461–467.

- Van Hooft, G., Benders, J., Delarue, A. and Procter, S. (2005). Teamworking: looking back and looking forward. *The International Journal of Human Resource Management*, 16(2), 167–173.
- Wageman, R. (2001). How leaders foster self-managing team effectiveness: Design choices versus hands-on coaching. *Organization Science*, 12(5), 559–577.
- Wiegmann, D. A., ElBardissi, A. W., Dearani, J. A., Daly, R. C. and Sundt III, T. M. (2007). Disruptions in surgical flow and their relationship to surgical errors: An exploratory investigation. *Surgery*, 142(5), 658–665.
- Wildman, J. L., Bedwell, W. L., Salas, E. and Smith-Jentsch, K. A. (2011). Performance measurement at work: A multilevel perspective. In S. Zedeck (Ed.), *Handbook of Industrial and Organizational Psychology: Building and Developing the Organization* (Vol. 1, pp. 303–342). Washington, DC: American Psychological Association.
- Wildman, J. L., Salas, E. and Scott, P. R. (2013). Measuring cognition in teams: A cross-domain review. *The Journal of the Human Factors and Ergonomics Society*, 56(5): 911–941. Advance online publication: doi: 10.1177/0018720813515907.
- Wildman, J. L., Thayer, A. L., Rosen, M. A., Salas, E., Mathieu, J. E. and Rayne, S. R. (2012). Task types and team-level attributes synthesis of team classification literature. *Human Resource Development Review*, 11(1), 97–129.





---

# 30 Human Reliability Assessment

*Barry Kirwan*

## CONTENTS

|                                                                               |     |
|-------------------------------------------------------------------------------|-----|
| Overview of Human Reliability Assessment .....                                | 791 |
| HRA Process .....                                                             | 794 |
| How HRA Works .....                                                           | 795 |
| Task Analysis.....                                                            | 795 |
| Human Error Identification .....                                              | 796 |
| Human Error Theory .....                                                      | 796 |
| Human Error Identification: Practice .....                                    | 798 |
| Human Error Quantification.....                                               | 804 |
| Evolution of Human Reliability Assessment.....                                | 804 |
| HRA, Ethics, Just Culture and the Resilience Perspective on Human Error ..... | 817 |
| Conclusion .....                                                              | 818 |
| References.....                                                               | 819 |

*Note to the reader:* The term *human error* is used throughout this chapter as a convenient label for something done or not done by the human operator, which is unintended by the operational system concept and can lead to increased safety risk. In Human reliability assessment (HRA), there is no connotation of blame or an action being wrong in a moral or pejorative sense. All actions are the product of the system via its design, environment, culture, training and other factors. HRA presumes that an error could happen to any of the operators, therefore blame is an inappropriate and misleading concept. This will be returned to later when discussing safety and Just Culture. HRA never ascribes blame to a particular person, since by definition another person would likely have had the same error. HRA uses the term because it is convenient, and even those who write with a focus on resilience and who dislike the very term human error often end up using it in their books, because it is difficult to find a workable substitute when discussing human performance within a risk management process.

## OVERVIEW OF HUMAN RELIABILITY ASSESSMENT

Human reliability assessment (HRA) is concerned with predicting the impacts of human performance and error on risk. HRA is a hybrid discipline, belonging simultaneously to the fields of risk assessment and reliability engineering, as well as ergonomics and human factors (E/HF). It seeks to deal with the difficult issue of human error, which can cause or contribute to accidents, but also with human recovery, which can save lives and prevent or reduce system losses. The human operator (a generic term used throughout this chapter) is the key ingredient in many industries and work systems, including nuclear power plants, chemical and oil and gas installations, aviation and air traffic management systems, rail networks, space missions, military systems and hospitals. These are all areas where HRA is currently in use or being developed to help avoid accidents leading to fatalities and unacceptable system losses. HRA usually fits into a risk (or reliability)

assessment, which is concerned with how a system can fail and/or lead to accidents involving loss of life or resources. Risk assessment asks several questions:

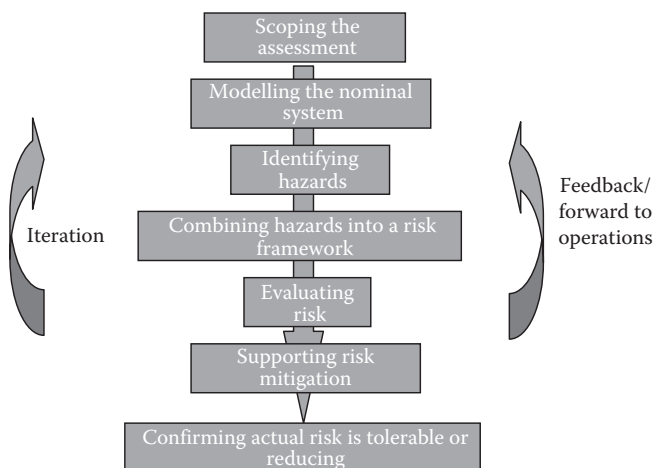
- What is the scope of the system under investigation?
- What can go wrong?
- How likely is it to go wrong?
- What will be the consequences?
- Is the risk (the product of the undesirable event's likelihood and consequences) acceptable?
- Can risk be reduced? Can error recovery be enhanced?

HRA therefore fits into a safety assessment or safety case aimed at answering the questions 'is the system safe enough?' and 'how can it be safer?' Such an approach is a safety risk management (SRM) approach, and people who step onto an airplane or live near a nuclear power plant expect that someone has calculated all the risks and put in required mitigations so that people will not be adversely affected or harmed in any way. It is important that HRA is seen in this context: it is part of a process to protect workers and public from harm, via identifying potential pathways to accidents, and then determining how to stop them from happening.

Figure 30.1 depicts the main stages of risk or safety assessment (Kirwan, 2006). HRA's role in risk assessment is to evaluate the human contribution to risk, which can be positive or negative. HRA's functions therefore reflect those of risk analysis. In particular, HRA usually attempts to answer the following questions:

- What is the human involvement in the system?
- What can go wrong?
- How likely is it to happen?
- What can be done to reduce error impact?

The first question is usually answered by the application of *task analysis* (Kirwan and Ainsworth, 1992), which defines the human's tasks, procedures to be followed and equipment required, among other things (see Chapter 6 of this book). For example, an air traffic controller might give the pilot of an aircraft an instruction to climb to a particular flight level using a radio-telephone. The task would be 'give instruction to pilot'. The second question is served by *human error analysis*, a range of techniques that aim to identify human errors that could cause or contribute to system failure or degradation.



**FIGURE 30.1** A generalised seven-stage safety assessment process.

A typical example is that the pilot mishears the flight level given by the controller and climbs to a higher level, one which is perhaps already occupied by another aircraft. Such an error can increase the risk of a mid-air collision. The third question is answered by using the techniques of *human error quantification*, which estimate how likely such an error is to happen, expressed as a probability. The *human error probability* (HEP) is ideally a statistical entity derived from observations of tasks and the errors that arise during their execution, that is,

$$\text{HEP} = \frac{\text{Number of errors observed}}{\text{Number of opportunities for error}} \quad (30.1)$$

In this example, based on research into this type of error, it may be expected to occur approximately three times in a 1000 opportunities. This sounds reasonably low, but there are many thousands of flights each day. To compensate for this, a number of recovery and mitigation factors are built into the system that the risk assessment can take account of; examples are as follows:

- The co-pilot in the aircraft may detect the error, since he or she will overhear the controller–pilot interaction.
- The pilot should read back the flight level to the controller, who can therefore detect the mishearing.
- The controller is watching the aircraft on a radar screen and may realise it is climbing beyond its cleared level.
- There is an alarm to warn the controller if two aircraft get too close.
- There is an independent alarm aboard the aircraft to alert the pilots and to tell them how to avoid the other aircraft.
- The other aircraft will be listening on the same frequency and may hear the instruction, whereupon the pilot will contact the controller immediately.
- There is sufficient space in the skies that chances of mid-air collision are still rare even if all these system factors fail to prevent the error and its recovery.

This all needs to be added up as a total risk picture to decide whether the system is safe enough. Currently the accident rate due to ATM errors in Europe is 1.55 accidents in a hundred million flight hours, equating to an accident every 2 years at prevailing traffic levels in European skies. In reality, at the time of writing, there has not been a fatal civil aviation accident due to ATM in Europe since 2002.

The impact of human error may therefore be deemed tolerable, because its effects can be absorbed or counteracted by other system defences, as in other industries. Alternatives such as full automation may seem like a solution, but experience shows that automation usually introduces its own problems, in particular a lack of recovery when it goes wrong, something that would be unacceptable when considering mass public transportation by air, for example. In aviation, full system-wide automation is unlikely before 2050.

At first sight, this example suggests that human reliability is quite good (e.g. three errors in a 1000 opportunities is not so bad), but when the sheer number of flights and controller–pilot interactions are considered, it appears problematic. In fact, upon detailed risk analysis, there are many recovery layers in the system, so the overall risk is acceptable. But for the mid-term future (e.g. 2030), where the number of flights is predicted to double or even triple in line with public demand for more air transportation, it may be necessary to introduce a new layer of defence. This could involve *human error reduction* measures such as sending an electronic message from the aircraft's onboard flight management system to the ground control system to compare what the controller input into the system with what the pilot input into the aircraft/cockpit-based system (a consistency check), and thereby alarming much earlier than current systems allow if an error occurs. Since the

first version of this chapter, this new system function has been successfully implemented in several countries in Europe, adding a new layer of defence.

This example shows that the human reliability for a task needs to be seen in the full system context to make meaningful decisions about risk. This tends to be true in all complex industrial domains. For this reason, HRA usually occurs within the framework of a system risk or reliability assessment (Cox and Tait, 1991). HRA analysts work closely with risk analysts, system designers and operational personnel. The HRA analyst's job may be described as being that of a detective investigating an accident, where the accident has not yet occurred. The overall aim is to ensure that the system is protected in the future both from known failure events and imaginable ones that have yet to occur.

## HRA PROCESS

A generic HRA process is shown in Table 30.1, from scoping the problem the HRA is being asked to address to documentation. Within the confines of this chapter, only three of these 10 steps will be covered: task analysis, error identification and quantification (with brief treatment of dependence). Most emphasis is placed on the error identification and quantification stages in HRA, the first to try to make the concept of human error clearer, and the second because it is most often seen as the quintessential element of HRA. These two steps generally dominate HRA outputs. For a fuller description of all 10 steps, see Kirwan (1994).

**TABLE 30.1**

### HRA Process (Using Aviation as an Example Context)

- *Problem scoping* – Deciding what is in the scope of the assessment – e.g. the principal human interactions and functions; what non-nominal events could occur (including adverse weather and traffic patterns); what failures, whether maintenance activities are included; etc. This 'context' is normally set by the safety assessor and HRA/human factors practitioner.
- *Task analysis* – The detailed analysis of the human contribution in the risk scenario – what the controllers/pilots should do, with what resources (displays, communication media, etc.).
- *Error identification* – What can go wrong – from simple errors of omission (e.g. fail to detect wrong readback) to more complex errors of commission (e.g. right instruction given to *wrong aircraft*).
- *Representation* – Modelling the human errors and recoveries in the risk model, whether this uses fault and event trees, Petri nets or some other modelling approach. This puts human errors and recoveries into the larger system context, alongside other system failures and defences. Representation enables an accurate view of how errors propagate through the system, and whether they will indeed ultimately affect system risk.
- *Dependence analysis* – Explicit quantitative accounting for dependence between different errors and error conditions.
- *Screening* – Conservative (pessimistic) error probabilities can be used along with sensitivity analysis to determine which human errors are 'risk sensitive' and which will have negligible impact on risk. The former can then be quantified in more depth, the latter ignored.
- *Quantification* – Derivation of HEPs for errors using a formal quantification method. Quantification may also develop uncertainty bounds or confidence levels for individual HEPs.
- *Evaluation* – Determination of the impact of human error on risk, and the most important human contributions to risk, within the risk framework or representation. This usually involves comparison of the total predicted risk against a target level of safety, or else future system risk will be contrasted with today's risk levels, to ensure that the future system will be at least as safe as today's.
- *Error reduction* – Based on the evaluation and the prior qualitative stages (task analysis, error identification and representation), human factors and/or safety requirements may be specified to reduce the human error probability and/or its impact on risk, to achieve a more acceptable risk level.
- *Documentation* – Documentation of the analyses, calculations, assumptions and recommendations to inform the designers, implementers and operators/managers, as well as to the appropriate regulatory authority.

The remainder of this chapter outlines the mechanics of HRA – what it looks like and how it works, showing examples of actual techniques. It then considers the validity of the central approach of human error quantification and finally discusses contemporary HRA issues including second-generation HRA techniques, safety culture and ethical issues in HRA.

## HOW HRA WORKS

This section gives examples of practical techniques and approaches of task analysis, error identification and quantification (including dependence aspects) and error reduction. For more general background information on HRA, see also Swain and Guttman (1983), Embrey et al. (1994) and USNRC (2005).

### TASK ANALYSIS

Task analysis is a set of methods for capturing and describing the human operator involvements in a system. The classic form of task analysis is hierarchical task analysis, an example of which is shown in Figure 30.2 (see Chapter 6 for further discussion). The figure is read top down and starts from a goal – in this case ‘launch lifeboat’ from an offshore oil and gas installation during an emergency. This goal is served by five subsidiary tasks which follow a plan requiring them to be carried out in order, for example check boat. This task requires two operations, shown in the diagram simply as A and B: ‘check maintenance valve is closed’ and ‘release retaining hook’. This simple type of hierarchical decomposition of a task is a useful way to define the operator’s role. Task 4 is decomposed in full detail to show the types of human operations that achieve the task ‘final checks’. This type of task analysis is a useful starting point for an HRA analyst and acts as a blueprint of what the human operator is supposed to do in a system.

However, it is important that the task analysis is not simply carried out from the analyst’s desk while reading procedures or working from documentation alone. The analyst needs to visit the actual operational centre or a good simulation of it, and have discussions with operators to see and discuss how operations are really carried out in practice.

Incidentally, this style of task analysis is also used by training departments to determine training needs, as well as by those operational groups that must write formal procedures for operators. Other formats of task analysis can be used to focus on, for example, team interactions under time constraints (using vertical and horizontal timeline analysis), operator interaction with a human–machine interface such as a computer screen (tabular task analysis) or distributed operations in different locations (link analysis). Task analysis can therefore describe the ‘nominal case’, that is, what the operator(s)

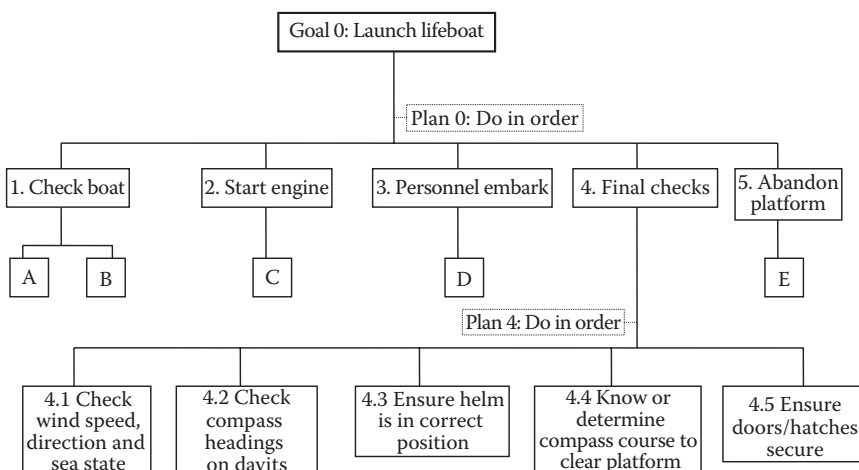


FIGURE 30.2 Example of HRA for launching a lifeboat.

should do (when carrying out an HRA at the design stage) and what the operators generally do (when carrying out an HRA for an existing system) – the rest of HRA then considers variability in human performance and what can go wrong from the system's viewpoint and how to increase reliability.

## HUMAN ERROR IDENTIFICATION

### Human Error Theory

Once task analysis has been completed, human error identification (HEI) can then be applied to consider what can go wrong. Human error has been defined as follows:

Any member of a set of human actions or activities that exceeds some limit of acceptability, i.e. an out-of-tolerance action where the limits of human performance are defined by the system. (Swain, 1989)

Such a definition gives rise to the following overt error types that should be considered during any human error analysis:

- *Omission error* – failing to carry out a required act
- *Action error* – failing to carry out a required act adequately:
  - Act performed without required precision, or with too much/little force
  - Act performed at the wrong time
  - Act performed in the wrong *sequence*
- *Extraneous act* – unrequired act performed (also called error of commission [EOC])

This classification tells us neatly what can happen, but gives little indication as to the causes or remedies for such errors. It also fails to capture the notion of different types of behaviour and error forms in terms of different causes and internal or cognitive failure mechanisms, and the important notion of intentionality of error, that is, that some errors are intentional (e.g. due to a misunderstanding of the situation) and others are not.

Although HRA is very applied in nature, the analyst must have an understanding of the theory of human error in order to be effective in identifying what can go wrong. Otherwise, error identification will be superficial and risk may be underestimated. To add more depth to the notion of human error, reference is usually made to four error categories (Reason, 1990):

- *Slip* – (attentional failure) where the wrong control is activated. For example, an error while dialling a number to make a phone call or the wrong message given, for example slips of the tongue.
- *Lapses* – (memory failure) where something is forgotten or a task step missed out, for example when making tea, filling the kettle but forgetting to switch it on.
- *Mistakes* – (wrong intention) where due to a misunderstanding or misconception, the operators do the wrong thing, for example in the Three Mile Island (TMI) nuclear power accident in 1979, operators erroneously believed the reactor pressure vessel had plenty of water, so they switched off the emergency supply when it tried to fill the emptying vessel, thus leading to reactor core damage. Mistakes imply that the intended task was executed, but it was the wrong task given the context.
- *Violations* – where the operators know what they are doing is not allowed, but they do it anyway, usually to make the job easier, for example stepping over a barrier around an automated assembly line to retrieve something dropped, without stopping the robot arm that is moving rapidly inside the barrier. Reasons for such a violation could include not wishing to lose valuable production time by stopping the process, and the risk of personal harm could be perceived as low. There is an element of risk-taking with violations, and operators committing a violation know that there can be negative repercussions for them personally (e.g. ranging from sanctions to criminal prosecution depending on the severity of the violation).



*Slips and lapses* are the most predictable error forms and are usually characterised as being simple errors of quality of performance or as temporary lapses in memory often due to distraction. In the lifeboat launching example, for example, a slip might concern pressing the wrong button to start the engine (there are several buttons on the console). But this would normally be recovered quickly (since the engine would not start – nothing would happen). A lapse might involve failing to ensure that all the hatches are secure, which could have more serious consequences when the boat enters the water. General good human factors practice can eradicate most of these types of errors, and good design will tend to mean that operators who commit them usually discover and correct them before they have adverse consequences on the system. This category however also includes slips and lapses in maintenance and testing activities, which may lead to *latent failures* (failures whose impact is delayed, and whose occurrence may be difficult to detect prior to an accident sequence). For example, an emergency back-up system may be taken offline for testing, then it may be tested and deemed working in good order, but the maintenance crew may omit the last critical step of placing it back online. This means that if it is suddenly needed in an emergency, it will be unavailable. This is why functional testing is recommended after maintenance activities have taken place. Latent failures are important because they can remove barriers to accidents, while letting the operators believe the barriers are perfectly healthy. In the lifeboat example, in the full task analysis which includes maintenance, there is a drain valve at the bottom of the boat, used in cleaning and maintenance activities. However, if it is left open and then an emergency evacuation occurs, the boat will slowly fill with water after it is launched.

*Mistakes* refer to the case wherein a human operator or team of operators might believe (due to a misconception) that they are acting correctly, and in so doing might cause problems and prevent automatic safety systems from stopping the accident progression. As noted earlier, such an error occurred in the TMI nuclear power plant accident in 1979 (which arguably was the launching point for HRA). In this event, following a reactor trip, the operators believed there was sufficient water in the reactor core, when in fact it was running out of coolant, and so when the automatic pumps engaged, the operators switched them off, leading to the core becoming uncovered and a partial core melt.

Usually, mistakes and misconceptions are more risk-significant than slips and lapses, because the latter are easier to identify and mitigate. To identify mistakes, the analyst must consider why an operator might do something they are not required to do, often at the worst possible time for them to do it. Because mistakes are due to a misunderstanding of the context, they often require deeper analysis than for slips and lapses. In part this has led to the development of second-generation HRA techniques, as outlined later in this chapter.

*Violations* represent a phenomenon that is not always explicitly considered in risk assessment, though one occurred most notably in the Chernobyl nuclear power plant accident in 1986. Rule violations inevitably involve some element of risk-taking or risk-recognition failure. There are two basic types (Reason, 1998): the first is the *routine* or *situational* violation, where the violation is seen as being of negligible risk and therefore the violation is seen as ‘acceptable’ or even a necessary pragmatic part of getting the job done. The second is the *extreme* violation, where the risk is largely understood as being real, as is the fact that it is a serious violation, with punitive penalties attached if discovered. The latter type of violation is believed to be rare in industry. An example of an extreme violation would be falsification of administrative control procedures, for example to say that a test had been carried out successfully, when it had not in fact occurred or had yielded a negative result. Such incidents as these are usually due to perceived ‘production pressure’ and may be seen as a way of keeping one’s job or ensuring the success of the company. Note that actual sabotage and terrorist acts are criminal in nature and are therefore outside the jurisdiction of HRA.

An EOC is one in which the operator does something that is incorrect and also *unrequired*. Examples are where a valve which should be locked open is found to be locked closed, or a computer-controlled sequence is initiated, or a safety system disabled, when none of these actions

are required, and in some cases when there has not even been a task demand associated with the equipment unit. Such errors can arise due to carrying out actions on the wrong components, or can be due to a misconception, or to a risk recognition failure. An EOC can therefore be due to a mistake or a violation, but even a simple slip could lead to an EOC, for example by activating a completely different (but adjacent) control rather than the one intended. EOCs are of concern for three reasons: firstly, they do appear to happen, even if rarely; secondly, they can have a large impact on system risk; and thirdly, they are very difficult and/or laborious to identify in the first place. This means that they may therefore be underestimated in terms of their contribution to risk, or may actually not even be represented in the total risk assessment. Their identification and quantification is difficult however and is returned to under the discussion of second-generation techniques in the latter part of this chapter.

### **Human Error Identification: Practice**

For a review of HEI techniques, see Kirwan (1998). This section outlines four straightforward and accessible approaches.

#### *Operational Experience (OPEX) Review/Critical Incident Technique (CIT)*

For an existing system, the first and most obvious way to identify errors is to review what errors have already led to incidents or accidents in the system under investigation. Most mature industries that are safety critical also have formal safety-related event reporting, recording and analysis systems, so that the organisation can learn from incidents and avoid their repetition or escalation towards more serious accidents. Such accidents usually involve fatalities and all the grievous losses associated with such tragic losses. In addition, large-scale accidents, particularly those that attract media attention (and most do), can also threaten the economic survival of the company. However, not all industries have mature human factors causal analysis systems, instead perhaps simply recording an event as ‘human error’ with no further analysis of why it occurred, and hence how to prevent recurrence. But if an organisation has a mature risk management culture, then it should have formal human error classification as part of its event reporting procedures and processes.

When carrying out an HRA for a system which already exists or has an equivalent referent system, it is very beneficial to interview operators (and also trainers, supervisors and incident investigators) to ask what types of events have happened. The analyst can simply ask if the interviewee (a single person or a group) know of any human error-related events that have occurred in the existing system in the past 10 years. While this is anecdotal, it can usually be corroborated by existing formal incident records, and such interviews often generate a richness of information, particularly about factors that can affect performance, and the state of the existing safety culture in the organisation. OPEX and CIT reviews and interviews mean that the HRA analysts have done their homework and will have a much clearer idea of the culture and contextual factors in the system being assessed. This will give them a feel for the overall state of the system, and what types of errors have been occurring, how well the procedures and recovery systems work, the effectiveness of training and the adequacy of the interface. OPEX and CIT are also useful means to ensure that the HRA does not omit scenarios or errors that have actually occurred in practice (unless they have been designed out since the events).

Two important caveats are first that the analyst has to beware of people’s particular biases (we all have them) such as a potential grudge against the company or the fact that everyone is a hero in their own story. Second, the analyst must avoid ‘leading the witness’ or putting words into their mouths based on the analyst’s own values and preconceptions.

At the least, reviewing incident and operational experience ensures that the analyst will be able to identify credible errors that have already happened and will gain a fuller understanding of the factors that affect human reliability in the system’s operational culture. CIT and OPEX set the scene for more focused error analyses.

### TRACER

In the early years of applied HRA (the early 1980s), errors were initially identified by risk assessors without any special techniques or processes. This changed as a result of the introduction of structured approaches for error identification, an early example being the systematic human error reduction and prediction approach (SHERPA: Embrey, 1986), which used a series of flowcharts and questions to help the risk or HRA analyst consider which errors could affect the system under investigation. A common approach used today is TRACER (technique for the retrospective analysis of cognitive errors: Shorrock and Kirwan, 2002), which is publicly available. It is a descendant of SHERPA as well as other techniques, and uses a series of taxonomies and flowcharts to identify error possibilities for a system. It also considers error detection and recovery. TRACER is a single-analyst HEI approach. It was developed to classify the causes of incidents, and then a further version (TRACER-Lite) was developed for predictive purposes. It requires first a task analysis and then applies a series of guidewords and other error-related taxonomies to identify both errors and (qualitative) error recovery likelihoods. The framework of TRACER is shown in Figure 30.3, and one of its main error taxonomies is shown in Table 30.2.

Note that in Table 30.2 there is a column called ‘internal error mechanism’; these guidewords are psychological in nature and attempt to describe likely psychological or cognitive mechanisms for the failure. These may be helpful later on when considering error reduction. As an example, if an action is predicted to be omitted, error reduction will be different if the most likely internal error mechanism was ‘distraction’ as opposed to ‘insufficient learning’. In the former case, changes to the work environment will need to occur to remove the distractions, whereas in the latter case, more comprehensive training is likely to be recommended. Application of TRACER is similar to using failure mode and effects analysis (FMEA: Cox and Tait, 1991) for hardware reliability assessment.

### Human HAZOP

Another approach to identifying errors is called human hazard and operability (HAZOP) study (Kirwan and Kennedy, 2001). This is a group-based approach which uses a set of guidewords of error types applied to a task analysis, with the group comprising a chairperson or group facilitator, a note-taker, a human factors expert, a safety expert, a system designer and operational experts. Table 30.3 shows typical guidewords used in a human HAZOP. Table 30.4 then shows an extract of

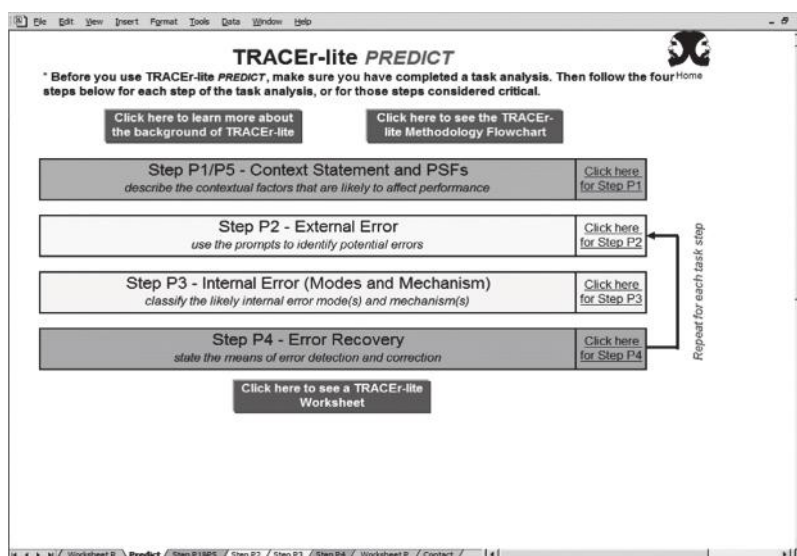


FIGURE 30.3 TRACER-lite overview.

**TABLE 30.2**  
**TRACER-Lite Internal Error Taxonomy (Guidewords)**

| Internal Error Mode        | Internal Error Mechanism                              |
|----------------------------|-------------------------------------------------------|
| <i>Perception</i>          |                                                       |
| Mishear                    | Expectation                                           |
| Mis-see                    | Confusion                                             |
| No detection (auditory)    | Discrimination failure                                |
| No detection (visual)      | Perceptual overload                                   |
|                            | Distraction/preoccupation                             |
| <i>Memory</i>              |                                                       |
| Forget action              | Confusion                                             |
| Forget information         | Memory overload                                       |
| Misrecall information      | Insufficient learning                                 |
|                            | Distraction/preoccupation                             |
| <i>Decision making</i>     |                                                       |
| Misprojection              | Misinterpretation                                     |
| Poor decision or poor plan | Failure to consider side effects or long-term effects |
| Late decision or late plan | Mind set/assumption                                   |
| No decision or no plan     | Knowledge problem                                     |
|                            | Decision overload                                     |
| <i>Action</i>              |                                                       |
| Selection error            | Variability                                           |
| Unclear information        | Confusion                                             |
| Incorrect information      | Intrusion                                             |
|                            | Distraction/preoccupation                             |
|                            | Other slip                                            |

**TABLE 30.3**  
**Typical Human HAZOP Guidewords**

|                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Basic guidewords</i>                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                    |
| <ul style="list-style-type: none"><li>• No action</li><li>• More action</li><li>• Less action</li><li>• Wrong action</li><li>• Part of action</li><li>• Extra action</li><li>• Other action</li><li>• More time</li><li>• Less time</li><li>• Out of sequence</li></ul> | <ul style="list-style-type: none"><li>• More information</li><li>• Less information</li><li>• No information</li><li>• Wrong information</li></ul>                                                                                 |
| <i>Additional concepts</i>                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                    |
|                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"><li>• Purpose/intention</li><li>• Clarity of task/scenario</li><li>• Training/procedures</li><li>• Abnormal conditions</li><li>• Maintenance</li><li>• Production pressure/risk-taking</li></ul> |

the type of discussion that can occur in a HAZOP, and Table 30.5 an example of how the results are recorded. The context of this example is a radar display for an air traffic controller, such as the one shown in Figure 30.4, where the ‘objects’ referred to in Table 30.4 are the individual aircraft labels. The ones shown here are basic ones, but those discussed in the example could allow the controller to see more information and carry out actions by using a mouse to place a cursor over the label.

The format in Table 30.5 can equally be used to represent the outputs from usage of TRACER or SHERPA or other HEI techniques. It is a useful means of documenting the error’s nature and

**TABLE 30.4**  
**Excerpt from HAZOP Discussion (of Electronic Strip System Design for Air Traffic Controllers)**

| HAZOP Group Member       | Discussion                                                                                                                                                                                                                        |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Human factors specialist | If these two objects overlap, could the controller operate on the wrong object, i.e. aiming a message for the one on top but actually communicating with the one that is now flying beneath the top one?                          |
| Designer 1               | Well, we expect the controller always to move the objects so that they won't overlap, before transmitting a command.                                                                                                              |
| Controller               | Well, there might not always be time, but as long as you're operating on the one on top, one coming underneath won't be selected without it being clear, will it?                                                                 |
| Designer 1               | Hmm. Well, it isn't impossible actually, depending on how long you leave the cursor without entering a command. Presumably that would be no more than a couple of seconds would it?                                               |
| Controller 2             | Not necessarily, I mean if I'm in the middle of something and then a higher priority call comes in, I'll leave the cursor there and then come back to it. It could be a while, up to a minute.                                    |
| Designer 2               | Right. Okay, we need to take another look at this, and implement some way of highlighting that the original target has been de-selected and must be re-acquired, otherwise the right message could be sent to the wrong aircraft. |
| Chairman                 | So, are we agreed then that we need an action on the designers here to ...?                                                                                                                                                       |

**TABLE 30.5**  
**Extract of Related ATM HAZOP Output**

| Function                          | Guide Word   | Cause                                                                         | Consequence                                               | Indication                 | System Defences                                                                                    | Human Recovery                                                             | Recommendations                                                                                                                                        |
|-----------------------------------|--------------|-------------------------------------------------------------------------------|-----------------------------------------------------------|----------------------------|----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| Highlight object (aircraft label) | No action.   | Another item preventing access to target.                                     | Difficulty in hooking target aircraft.                    | No highlighting of target. | None.                                                                                              | Drag blocking object out of way as a strategic management of screen items. | Design objects to 'roll around' each other; use height filtering, flip system to move between object on top and the one beneath, highlight background. |
|                                   | Other action | Clustering results in different aircraft being highlighted instead of target. | Instruction may be given to wrong aircraft on the system. | As above.                  | Highlighting is colour coded to indicate direction of travel; call sign is displayed on all menus. | As above.                                                                  | As above.                                                                                                                                              |

impacts on the system, including barriers that should prevent the error or its effects, and a useful medium with which to identify preliminary error reduction or mitigation measures.

HAZOP is a risk and reliability approach, and human HAZOP is simply an adaptation to HRA. However, in the few validations (Kirwan, 1992) that have taken place of error identification techniques, human HAZOP has been shown to be very effective, particularly in identifying credible errors in the areas of violations.



**FIGURE 30.4** Screenshot of an air traffic controller's radar display. The blocks show the position of the aircraft; the solid lines show aircraft direction. Other information includes the aircraft 'callsign' and its current height (300 = 30,000 ft above sea level).

HAZOP has three significant advantages. Firstly, when the safety assessor is trying to identify hazards for a future system, one where there is, as yet, no mature operational concept or procedures, this can be very difficult. There are simply too many open questions to make progress. However, in a HAZOP where experienced operators are present, these experts can usually fill in the gaps, interpolating between current practice and future practice of how the system should operate. This means that preliminary hazards can be identified at a time when the design can still be changed significantly if serious hazards emerge. The second and related advantage is that the composition of the HAZOP group, comprising project design and controller reviewers, often lends itself to promptly identifying design solutions that will increase safety. Once the hazard is identified, the participants will often determine an immediate solution via a system change or procedural modification (as in Table 30.4). The third advantage is that such a process involves the project design and operational personnel in safety discussions. If they really believe a hazard is incredible, they will say so, but if not, there is acceptance that such a hazard must be addressed. The approach therefore gains 'buy-in' or ownership from the project team. They 'own' the results as much as the safety assessors, rather than being given an 'external' safety assessment that is seen as a 'fait accompli', which they have to accept only on the basis that they trust that the safety people know what they are doing.

The disadvantages of HAZOP are the resources requirements, since around six domain experts are needed to run a HAZOP and progress can sometimes seem slow. Also, experts may be biased not to identify failures (e.g. if they represent the project or development team). The ideal approach is therefore to utilise an analytic approach like TRACER as well as a human HAZOP approach – this is the most comprehensive formula for identifying all likely errors. If TRACER is carried out first, then its results can be used judiciously to steer the resources of a HAZOP team more effectively, for example ignoring already identified 'minor' errors with little impact and focusing on the more serious potential errors and risks of violations, etc.

#### *JHEDI*

A final checklist approach is shown in Table 30.6, from a nuclear power technique known as JHEDI (justification of human error data information: Kirwan, 1997a). This straightforward technique can be used by a single assessor as a preliminary screening error identification tool.

The overall approach to HEI can be task-driven or system-risk-driven, and the choice depends on the system being investigated and the role of the operator. For air traffic control or surgery, for example, the operator (controller; surgeon) is executing the tasks manually and in real time, with their skill and tools as their main defence. A task analysis of such skilled performance and associated error analysis will yield many possible errors that could occur. Such an approach would be relevant if, for example, evaluating the



**TABLE 30.6****JHEDI Human Error Analysis Questions**

1. Could a misdiagnosis or misinterpretation occur, possibly due to a misperception of signals, or expectations, the complexity of the situation itself, or a failure to take note of special circumstances?
2. Could a shortcut or even rule violation occur due to production pressure or a failure to be aware of the hazard?
3. Could a previous (latent) maintenance or calibration error lead to difficulties or errors in responding to the scenario?
4. Could a check fail, be omitted or occur on the wrong system, or simply be the wrong kind of check?
5. Could an information–communication error occur?
6. Could a crew-functioning error occur, such that, for example, it is not clear who is carrying out the task?
7. Could a required act be omitted because it is at the end of the task, because of a lapse of memory or distraction, or because the operator failed to respond (e.g. due to his or her workload)?
8. Could the task be carried out in the wrong sequence, or a sub-task be repeated, or the operator lose his or her place in the procedures?
9. Could an action or check be carried out too early or too late?
10. Could an action be carried out with insufficient precision or force, or just inadequately?
11. Could a *non*-required action be carried out by operating on the wrong object or in the wrong location, or by carrying out an incorrect but more familiar task, or by selecting or acting on the wrong information ?
12. Could any other type of error occur?

risks and benefits associated with the use of a new tool by a controller or surgeon. This type of analysis is bottom-up; at the start, the risks that can arise from such operations are not necessarily known.

For a nuclear power plant, on the other hand, many tasks are automated or else have several layers of safeguards (defence in depth). A bottom-up task-driven analysis here would be highly resource-intensive and would result in many errors that had little or no consequences. Instead, usually the system risk analysis identifies those failure scenarios where the operator can make a difference and has an essential safety role to play (e.g. if automatic shutdown fails following an emergency event). These are called ‘claims on the operator’. Nuclear power plant risk assessments often focus on post-trip scenarios, where an event has occurred pushing the reactor system into the danger zone, and operator action is required to bring it safely under control (a reactor trip means the control rods are inserted into the reactor core – it is no longer at power, but can heat up if not cooled, leading to core breach and release of radioactivity into the environment). This identification of critical tasks based on the preliminary risk analysis narrows the search space for error identification, enabling HRA analyst resources to focus on the most risky tasks.

It is not always so clear cut, of course. For example, a nuclear power plant risk assessment may consider maintenance tasks in case they lead to the event that pushes the reactor into the danger zone, in which case this part of the HRA will analyse a broader range of tasks and roles, and will be more ‘bottom-up’ in nature, looking for risks that can arise from human actions and inactions, rather than starting from a clear set of risk-based scenarios.

Even where the HRA is ‘system-risk-driven’, it can be useful to carry out a degree of detailed task analysis and error identification (e.g. HAZOP) to add a layer of confidence that the system risk analysis has not missed anything critical, especially where errors of commission and latent errors are concerned.

### *‘Drift into Danger’ and Sabotage*

There are two areas of human involvement that HRA does not normally address. The first is the so-called drift into danger phenomenon, where safety standards have slowly slipped over a period of years, and unsafe practices have become normal. This is not always detectable by the analyst unless working together with an experienced operator from a different system where such drift has not occurred (such an operator would notice it very quickly). HAZOPs can also sometimes detect drift, often if there are both experienced and younger operators present. Drift into danger can sometimes be detected by safety culture surveys (see later).

As discussed later, sabotage is outside the remit of HRA, because it involves wilful harm, and would require a different model of human performance. Sabotage is in the domain of security rather than safety. However, HEI techniques could be used to explore vulnerabilities in the system. Probably such techniques would need to be extended, but they could serve as a useful starting point.

Once error identification is carried out, there are two paths open to the HRA analyst or risk analyst. The first is where only a quantitative approach is being adopted, that is, the intent is to examine the vulnerability of the system and seek ways to improve it. In such a case, it is likely that the errors identified (by whatever means) will be ranked or categorised in order of importance, and the assessors will move directly to the step of error reduction or mitigation.

The second path is taken when the HRA is part of a quantitative risk assessment approach. In such cases, the human errors will usually be represented in logic trees such as fault and event trees – these aspects are more properly dealt with in other chapters. If this pathway is being pursued, the next step requires human error quantification, that is, determining the probability of the human error occurring (where possible together with its uncertainty bounds and probability distributional characteristics). This step is usually seen as the major part of HRA.

## HUMAN ERROR QUANTIFICATION

Because human error quantification has been the main part of HRA for over four decades, a brief history of the development is given, before moving on to show how typical techniques work.

### Evolution of Human Reliability Assessment

HRA began in the 1960s at the same time that risk assessment practices were also developing, in the area of missile system development, most notably in various governmental-sponsored laboratories (e.g. Sandia National Laboratories) in the United States. For the first two decades, however, HRA was more of a research area than a practical discipline. It ‘came of age’, however, shortly after the US nuclear power plant TMI accident happened in 1979. This accident, a partial melting of the nuclear reactor core, was seen as a major demonstration of the vulnerability of systems to human error, and so an approach to predict and prevent such errors was urgently required. This led to the first practical HRA technique, the technique for human error rate prediction (THERP: Swain and Guttman, *op. cit.*), which is still in use today.

#### *THERP*

Application of THERP entails carrying out a task analysis of the human involvements in the system, defining human event sequences that need to be carried out to assure safety, and then quantifying these sequences using a database of human error probabilities. THERP uses a database of human error probabilities as its quantification ‘engine’. This database contains error probabilities for such tasks as reading a display, operating a joystick, responding to an alarm, etc. Some examples of THERP HEPs are shown in Table 30.7.

Table 30.7 suggests that human reliability is poor in the first 20 min after the onset of a system abnormality or fault condition. This unreliability is also because nuclear power plants are complex and so some basic diagnosis must take place to determine the fault according to its annunciated symptoms.

THERP recommends consideration of the effects of a large number of factors that can influence human performance (time pressure being one, in the example in Table 30.7), called performance shaping factors (PSF). These factors can range from ‘external’ factors, such as lighting, work hours and manning parameters, to more ‘internal’ factors, such as motivation, experience and stress. The ‘error factor’ in Table 30.7 associated with each nominal HEP is then used as a modification factor, for example if there was significant stress on the operators, an HEP might be multiplied by a factor of 10.

THERP was also the first technique (and one of the few) to explicitly consider *dependence*. Dependence is when the success or failure of one action is related to that of a previous action or task.

**TABLE 30.7**

**Extract from THERP's Chapter 20 [6] on HEP Data: HEPs for Failing to Diagnose an Annunciated (Alarmed) Nuclear Power Abnormal Event**

| Item | T (Minutes after Onset) | Median Joint HEP for Diagnosis of a Single, or the First, Event | Error Factor |
|------|-------------------------|-----------------------------------------------------------------|--------------|
| 1    | 1                       | 1.0                                                             | —            |
| 2    | 10                      | 0.5                                                             | 5            |
| 3    | 20                      | 0.1                                                             | 10           |
| 4    | 30                      | 0.01                                                            | 10           |
| 5    | 60                      | 0.001                                                           | 10           |
| 6    | 1 day                   | 0.0001                                                          | 30           |

For example, if an operator is calibrating a set of gas detectors and makes a mistake in one, for example, because he has the wrong value in mind, then he is likely to miscalibrate all of them. Another example of dependence is when someone is checking another's work or results – there is usually a certain degree of trust so that the check may not be as thorough as one would hope for. In such cases, adding a third or even fourth 'checker' into the equation will not necessarily increase reliability either, due to a natural dependence.

THERP models dependence in five categories, complete, high, moderate, low and zero dependence, and modifies the second (dependent) HEP accordingly using a specific formula. The analyst judges dependence according to whether the actions under consideration are being carried out by the same or different people or teams, whether they are using the same equipment in the same location and whether they are in the same timeframe or not. A useful guide to dependence modelling and countermeasures is given in Greenstreet Berman (2001).

### *Expert Judgement Methods*

Although THERP was immediately successful in the early 1980s, there were a number of criticisms of the approach, in terms of its granularity of task description and associated human error probabilities. It was seen by some as too 'decompositional' in nature. Additionally, its database origins have never been published, and there were concerns from the human factors community about its 'broad brush' treatment of psychological and human factors aspects, as well as its significant resource requirements when it came to analysing a complex system such as a nuclear power plant.

By 1984, there was therefore a swing to a new range of methods based on *expert judgement*. Risk assessment at the time (specifically probabilistic safety assessment [PSA]), whose champion was (and still is) the nuclear power industry, was used to applying formal expert judgement techniques to deal with highly unlikely events (e.g. earthquakes and other external events). These techniques were therefore adapted to HRA. In particular, two techniques emerged: absolute probability judgement (Seaver and Stillwell, 1983), in which experts directly assessed HEPs on a logarithmic scale from 1.0 to  $10^{-6}$ , and the success likelihood index method (SLIM: Embrey et al., 1984). The main difference with this technique was that it allowed detailed consideration of PSF in the HEP calculation process. Typical performance shaping factors used by a number of HRA techniques are the following:

- Unfamiliarity of the task (to the human operator)
- Time pressure or stress
- Training adequacy
- Procedures adequacy
- Adequacy of the human–machine interface
- Task complexity

Expert judgement approaches are useful for considering difficult quantification assignments, for which perhaps there are no clear data available, or where stakeholders themselves want to be sure that the numbers are judged realistic by operational experts. In this sense, expert judgement techniques have some of the advantages of HAZOPs, since they elicit ownership and involvement from the participants, as well as maximising expertise input by having operational experts involved. Conversely, if the operational experts are not very experienced themselves (a recommended minimum operational experience level is 10 years, and to still be an active operator), or if the system being assessed is way outside their own experience range, then there is the danger of gaining results that are not accurate or are invalid. There is also the danger of biases affecting the judgement process, whether these are memory based (e.g. giving disproportionate weight to a rare event and ignoring other more representative ones), motivational (e.g. wanting to see a new system built, so underestimating the HEPs) or cognitive (e.g. finding it hard to deal with very low probabilities such as one error in a hundred thousand).

#### *HEART, NARA, CARA, RARA and SPAR-H*

One further technique worthy of note developed in the mid 1980s was the human error assessment and reduction technique (HEART; Williams, 1988). This technique had a much smaller and more generic database of HEPs than THERP, which made it more flexible and quicker to use, and had PSF called error producing conditions (EPCs), each of which had a maximum effect (e.g. from a factor of 19 to a factor of 1.2) on the HEPs. HEART was based on a review of the human performance literature (field studies and experiments) and so the relative strengths of the different factors that can affect human performance had some credibility with the human factors and reliability engineering domains. HEART is still in use today in a number of industries as it is very resource-efficient and has recently been redeveloped as a specific nuclear power plant HRA technique called NARA (nuclear action reliability assessment; Kirwan et al., 2004). As an example of how these techniques work, Tables 30.8 and 30.9 show the generic task types (GTT) for NARA and the error-producing conditions.

The HRA analyst assessing a nuclear power plant therefore, having identified a critical human action, for example, responding to an alarm, must use Table 30.8 to pick the appropriate (best fit) GTT. Assuming it is not a complex pattern, C1 would be selected, with a 'nominal' HEP of 0.0004 or 4 errors per 10,000 opportunities (incidentally, this particular GTT is based on observed error rates with experienced UK nuclear power plant operators). The assessor must then (based on prior task analysis and interviews with existing operators or prospective ones if the system is not yet built) consider if there are any human factors that could increase human error likelihood. As with most HRA techniques, NARA pre-identifies the human factors that are most important for HRA consideration, along with 'weighting factors' to determine how large the impact on the HEP will be.

NARA has some additional functionality: it has an extended time factor module, for considering scenarios where there is significant additional time to retry an action which did not work at first. This is perhaps a British nuclear power plant phenomenon, since some of the gas-cooled reactor plants (as opposed to light water reactors, the most common type of nuclear power plant worldwide) have long timescales to respond (e.g. up to 24 h). Additionally, there is guidance on how to assess the strength of EPCs for each scenario, and also guidance on avoiding 'double-counting' of negative impacts from GTTs and EPCs.

An ATM version and a rail version have also been developed, called CARA (controller action reliability assessment; Kirwan et al., 2011) and RARA (railway action reliability assessment; Gibson et al., 2012), respectively. The associated GTTs are shown in Tables 30.10 and 30.11, respectively (both techniques are currently freely available online). Figure 30.5 shows the (information processing-based) model of cognitive human performance underlying CARA (the GTTs reflect this model of air traffic controller performance).

**TABLE 30.8**  
**NARA Generic Task Type HEPs**

| GTT ID                                                                   | NARA GTT Description                                                                                                                                                                                                               | GTT HEP |
|--------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| <i>Task execution</i>                                                    |                                                                                                                                                                                                                                    |         |
| A1                                                                       | Carry out a simple single manual action with feedback. Skill based and therefore not necessarily with procedures. Independent of location.                                                                                         | 0.005   |
| A2                                                                       | Start or reconfigure a system from the main control room following procedures, with feedback. The task may include some decision making, if supported by clearly set out procedures.                                               | 0.001   |
| A3                                                                       | Start or reconfigure a system from a local control panel following procedures, with feedback.                                                                                                                                      | 0.002   |
| A4                                                                       | Completely familiar, well-designed highly practised, routine task performed by highly motivated, highly trained and experienced person, totally aware of implications of failure, with time to correct potential errors.           | 0.0001  |
| <i>Ensuring correct plant status and availability of plant resources</i> |                                                                                                                                                                                                                                    |         |
| B1                                                                       | Routine check of plant status.                                                                                                                                                                                                     | 0.02    |
| B2                                                                       | Restore a single train of a system to correct operational status after a test, following procedures.                                                                                                                               | 0.004   |
| B3                                                                       | Set system status as part of routine operations using strict administratively controlled procedures, e.g. top up tank to correct level.                                                                                            | 0.0005  |
| B4                                                                       | Calibrate plant equipment using procedures, e.g. adjust set-point.                                                                                                                                                                 | 0.003   |
| B5                                                                       | Carry out analysis.                                                                                                                                                                                                                | 0.03    |
| <i>Alarm/indication response</i>                                         |                                                                                                                                                                                                                                    |         |
| C1                                                                       | Simple response to alarms/indications providing clear indication of situation (simple diagnosis required). Response might be direct execution of simple actions or initiating other actions separately assessed.                   | 0.0004  |
| C2                                                                       | Identification and response to situation requiring detection of alarms/indications embedded within other alarms/indications. Response might be direct execution of simple actions or initiating other actions separately assessed. | 0.04    |
| C3                                                                       | Identification and response to a situation requiring interpretation of complex pattern of alarms/indications. Response might be direct execution of simple actions or initiating other actions separately assessed.                | 0.2     |
| <i>Communication</i>                                                     |                                                                                                                                                                                                                                    |         |
| D1                                                                       | Verbal communication of safety-critical data.                                                                                                                                                                                      | 0.006   |

The DRA is a reminder device in the train cab that, once activated, presents a red light to the driver and prevents the driver from taking power. The driver is expected to set the device at a station where the red aspect of the next signal is visible. The aim of the DRA is to reduce the likelihood of a driver starting against a signal at danger in error.

For the scenario which was assessed using railway action reliability assessment, the driver comes to a stand at a platform, with the next signal visible from the platform and displaying a red signal aspect. With a red signal aspect visible, the driver should set the DRA (driver's reminder appliance) by pushing in a DRA button. The specific error which was assessed was that when the driver arrives at the platform the driver continues with other station duties without setting the DRA. The scenario considered that the station was busy with passengers.

Setting the DRA is a routine action which can occur a number of times per hour and is therefore familiar and skill based. However, it is also a task open to simple confusion as the driver will be required to press the DRA at some station stops and not at others. This task was reviewed against the set of GTTs presented in Table 30.11 and the GTT R4 'skill-based tasks (manual, visual or communication) when there is some opportunity for confusion' was selected as the best match. Note that violations can be associated with DRA operation, but this is outside the scope of assessments using railway action reliability assessment.

**TABLE 30.9**  
**NARA Error Producing Conditions**

| NARA No. | EPC                                                                                                                            | Maximum Affect |
|----------|--------------------------------------------------------------------------------------------------------------------------------|----------------|
| 1        | Poor, ambiguous or ill-matched system feedback                                                                                 | 30             |
| 2        | Unfamiliarity, i.e. a potentially important situation which only occurs infrequently or is novel                               | 17             |
| 3        | A need to unlearn a technique and apply one which requires the application of an opposing philosophy                           | 13             |
| 4        | Time pressure                                                                                                                  | 11             |
| 5        | Low signal-to-noise ratio                                                                                                      | 10             |
| 6        | A conflict between immediate and long-term objectives                                                                          | 9              |
| 7        | No obvious means of reversing an unintended action, i.e. carrying out a complex task where the system is unforgiving of errors | 8              |
| 8        | A means of suppressing or over-riding information or features which is too easily accessible                                   | 7              |
| 9        | Operator inexperience                                                                                                          | 7              |
| 10       | Cognitive overload, particularly one caused by simultaneous presentation of non-redundant information                          | 6              |
| 11       | No obvious way of keeping track of progress during an activity                                                                 | 6              |
| 12       | Shortfalls in the quality of information conveyed by procedures                                                                | 5              |
| 13       | Difficulties caused by poor shift handover practices and/or team coordination problems or friction between team members        | 4              |
| 14       | An incentive to use other more dangerous procedures to achieve long-term objectives                                            | 4              |
| 15       | Poor environment                                                                                                               | 3              |
| 16       | High emotional stress and effects of ill health                                                                                | 2              |
| 17       | Low workforce morale or adverse organisational environment                                                                     | 1.5            |
| 18       | Operator underload/boredom                                                                                                     | 1.2            |

The scenario assessed included the potential for additional workload because the platform is busy with passengers. Also, this scenario was assessed for driver only operation, where the driver would have responsibility for platform duties, which could lead more directly to distraction type effects. These factors were reviewed against the railway action reliability assessment EPCs presented in Gibson et al. (2012). The EPC 'A conflict between immediate and long-term objectives' was applied to represent the conflict between the routine DRA task and platform duties for driver-only operation. A busy platform is identified to be a routine occurrence and an issue which drivers are trained to manage. The experience and training of the driver in managing these situations were assessed to lead to an assessed proportion of affect of 0.5. *Time pressure* (T2) may also be an EPC which comes into play at station stops, but this has not been included as it was not included in the scope for this scenario.

Based on the railway action reliability assessment approach, this led to the following components to produce the quantitative calculation. The full details of the calculation method are presented in Gibson et al. (2012).

As the GTT selection was '*Skill-based tasks* (manual, visual or communication) when there is some opportunity for confusion', this provides an initial human error probability of 0.003.

As the EPC T7 '*A conflict between immediate and long-term objectives*' was selected with an assessed proportion of affect 0.5, this modifies the human error probability, increasing it to a value of **0.005**.

It should be noted that in line with the proposed approach for railway action reliability assessment, these assessments have been reviewed and modified based on reviews with technical and human factors specialists.

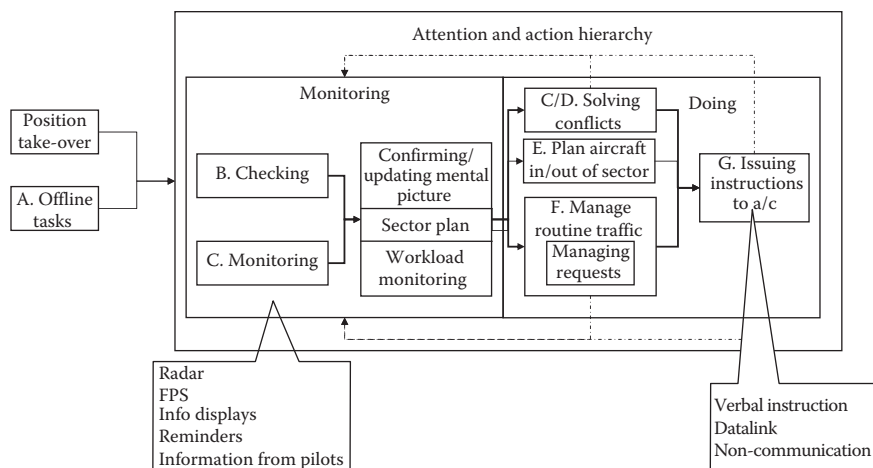


**TABLE 30.10**  
**CARA GTTs**

| Task Context                                         | Generic Task Type                                                                                                            | Examples of Task Errors                                                                                                                                                                                 | HEP    | Uncertainty Bounds |
|------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|--------------------|
| A. Offline tasks                                     | A. Offline tasks                                                                                                             | Fail to check, or fail to extract key information from, NOTAMS or planned maintenance notices.                                                                                                          | 0.02   | —                  |
| B. Checking                                          | B1. Active search of radar or FPS, assuming some confusable information on display                                           | Misread numeric value in datablock (e.g. flight level, callsign, etc.).                                                                                                                                 | 0.005  | 0.002–0.02         |
|                                                      | B2. Detect unannounced change in visual alert                                                                                | Fail to notice that highlighted aircraft changes to low-lighted.                                                                                                                                        | 0.13   | 0.05–0.3           |
|                                                      | B3. Respond to unique and trusted audible and visual indication                                                              | Fail to respond to robust and standardised short-term conflict alert.                                                                                                                                   | 0.0004 | —                  |
| C. Monitoring for conflicts or unanticipated changes | C1. Identify and resolve routine conflict                                                                                    | Fail to identify conflict between two converging aircraft, where loss of separation will occur within 2–5 min (e.g. prior to STCA or TCAS occurrence). Note that the resolution element is not complex. | 0.02   | —                  |
|                                                      | C2. Identify unanticipated change in radar display                                                                           | Fail to identify change in digital flight level due to aircraft deviation or corruption of datablock value (e.g. callsign code conversion error).                                                       | 0.3    | 0.2–0.5            |
| D. Solving conflicts                                 | D1. Resolve impending loss of separation                                                                                     | Fail to resolve conflict which requires one instruction in non-complex situation. Several solutions are possible. STCA may have already occurred.                                                       | 0.02   | Holding value      |
|                                                      | D2. Complex and time pressured conflict solution                                                                             | Fail to resolve conflict which has been identified late and is constrained by other traffic. STCA may have already occurred.                                                                            | 0.19   | 0.09–0.39          |
| E. Plan aircraft in/out of sector                    | E. Plan aircraft in/out of sector                                                                                            | Fail to plan conflict-free traffic in to or out of sector.<br>Fail to warn executive controller of impending conflict situation.                                                                        | 0.01   | Holding value      |
| F. Manage routine traffic                            | F. Routine element of sector management (e.g. rule-based selection of routine plan for an aircraft or omission of clearance) | Fail to appropriately sequence aircraft in TMA or for take-off/landing.                                                                                                                                 | 0.003  | Holding value      |
| G. Issuing instructions                              | G1. Verbal slips                                                                                                             | State flight level 110 instead of flight level 120.<br>State ‘turn left’ rather than ‘turn right’.<br>State ‘in your 10 o’clock’ rather than ‘in your 2 o’clock’.                                       | 0.002  | 0.001–0.003        |
|                                                      | G2. Physical slips (two simple choices)                                                                                      | Error in menu selection using a mouse.<br>Mis-key number using a numeric keypad.                                                                                                                        | 0.002  | 0.0008–0.004       |
| M. Technical and support tasks                       | M. Routine maintenance task                                                                                                  | Fail to successfully complete instrumented landing system (ILS) return to service checks.                                                                                                               | 0.004  | 0.0009–0.01        |

**TABLE 30.11**  
**RARA (Railway Action Reliability Assessment) GTTs**

| Area                                     | Generic Task Type                                                                                                                | HEP     |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|---------|
| More automated and skill-based processes | R1. Respond correctly to system command even when there is an automated system providing accurate interpretation of system state | 0.00002 |
|                                          | R2. Completely familiar, well-designed, highly practised task which is routine                                                   | 0.0004  |
|                                          | R3. Simple response to a dedicated alarm and execution of actions covered in procedures                                          | 0.0004  |
|                                          | R4. Skill-based tasks (manual, visual or communication) when there is some opportunity for confusion                             | 0.003   |
|                                          | R5. Fairly simple task performed rapidly or given insufficient or inadequate attention                                           | 0.09    |
| More effortful and rule-based processes  | R6. Restore or shift a system to original or new state, following procedures with some checking                                  | 0.003   |
|                                          | R7. Identification of situation requiring interpretation of alarm/indication patterns                                            | 0.07    |
| Thinking outside procedures              | R8. Complex task requiring a high level of understanding and skill                                                               | 0.16    |



**FIGURE 30.5** Human cognitive performance model for CARA.

The final calculated value using railway action reliability assessment was that the likelihood per operation that the driver fails to set the DRA for the scenario was 0.005. This equates to a prediction of 1 error occurring every 200 times the DRA is required to be operated. This value, before being used in assessment, would be subject to review against known data on performance for the task and ideally be supported with review by drivers experienced in the task.

In order to show the detail of the HEP derivation, here is a worked example broken down using NARA, wherein there is a relatively simple single alarm to respond to. Let us assume however that in this case the assessor finds that the human-machine interface for displaying alarms is not as good as it should be and that the alarm condition is rare and will be relatively unfamiliar to the operator. The assessor would then select EPCs 1 (poor or ill-matched system feedback) and 2 (unfamiliarity), with maximum effects on the HEP of 30 and 17, respectively. The assessor then has to make a judgement on how unfamiliar the condition is, and how poor the interface is, and this requires assessor training and experience with the technique itself (in NARA descriptive 'anchor points' for each EPC also assist the assessor). Let us suppose the analyst assesses that rather than using the full impact

of these EPCs, 0.2 of the interface maximum effect and 0.5 of unfamiliarity should be applied. This is called (in HEART as well as NARA) the assessed proportion of affect (APOA). The values chosen are then substituted into the following generic NARA quantification equation:

$$\text{HEP} = \text{GTT} \times \{[(\text{EPC}_1 - 1) \times \text{APOA}_1 + 1] \times \dots \times [(\text{EPC}_n - 1) \times \text{APOA}_n + 1]\} \quad (30.2)$$

Inputting the chosen GTT and EPC information gives the following result:

$$\begin{aligned} \text{HEP} &= 0.0004 \times \left\{ [(30 - 1) \times 0.2 + 1] \times [(17 - 1) \times 0.5 + 1] \right\} \\ &= 0.0004 \times [(29 \times 0.2) + 1] \times [(16 \times 0.5) + 1] \\ &= 0.0004 \times 6.8 \times 9 = 0.025 \end{aligned}$$

Thus the final HEP is approximately 1 in 40 opportunities or system demands. This value would then be put into the appropriate fault or event tree. NARA also assesses dependence using a version of the original THERP dependence model.

The form of Equation 30.2 prevents inadvertent reduction of the HEP by selecting a low APOA with a low maximum effect EPC. For example, without this adjustment, if an APOA of 0.1 was selected for EPC 13 (poor environment), with the same GTT (C1), this would lead to a lower HEP than the original GTT:

$$\text{HEP: } 0.0004 \times 0.1 \times 3 = 0.00012$$

Clearly this is not the intended outcome, so Equation 30.2 avoids this problem. One other point to mention is that in some cases the multiplication factors push the HEP beyond the bounds of probability, for example producing an outcome value greater than unity (1.0), in which case the analyst must simply truncate the value at 1.0. As mentioned already, any HRA technique requires training and experience, and data sets (scenarios, factors and actual [known] HEPs) are available with which HRA analysts can train themselves and 'calibrate' their approach.

The example also highlights another useful function of this type of analysis. Of the two EPCs, which is most important if error reduction is required? In this particular example, the problem of unfamiliarity has a larger impact on the human contribution to risk, so if error reduction is required, this would be the first port of call (e.g. via refresher training every 6 months). This is a useful function, and HRA can help human factors prioritise issues – it may be far more effective and cheaper, for example, to implement a refresher training regimen than provide an upgrade to the whole alarm system.

A similarly flexible and quick HRA technique developed in the United States and applied on over 70 HRAs to date is the standardised plant analysis risk model HRA (SPAR-H) technique (Gertman et al., 2005). This uses an HEP calculation process and a small number of performance-shaping factors with associated quantitative multipliers to quantify HEPs for diagnosis and subsequent actions in nuclear power plants. It also contains guidance for dependence and uncertainty calculations. SPAR-H is widely used in the United States and is to an extent based on data, like the HEART and more recent NARA family of techniques.

### *Human Error Data*

Human error probabilities would ideally be based on empirical data and stored in databases much as for reliability data in some industries. In fact, although there have been numerous human error databases over the years (e.g. Topmiller et al., 1984; Gertman and Blackman, 1994), few have survived. One current human error database that is still growing is the computerised operator reliability and error database (CORE-DATA, Kirwan et al., 1997). This database has some high-quality data from nine separate industrial sectors and has been the basis for development of NARA, CARA (see Kirwan and Gibson, 2007) and RARA. Some example data from CORE-DATA are shown in Table 30.12.

**TABLE 30.12**  
**Examples of HEPs from the CORE-DATA Database**

| Study                        | Description of Error                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | HEP     | Data Source       |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|-------------------|
| Nuclear (confidential)       | During a shift the transport department brought a chemical load to the compound after permission had been arranged between two supervisors, but the correct paperwork did not arrive with the chemicals. Consequently, this led to two cans of highly enriched chemical solution being processed instead of six cans of low enriched chemical.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 0.0027  | Real data         |
| Manufacturing (confidential) | A component has a different profile machined on each end. The operator inadvertently machines the aft end profile on the forward end.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0.00039 | Hardcopy data     |
| Chemical [16]                | In a simulated chemical process plant, a trainee in a diagnostic task is faced with a panel of 33, instruments eight of which they were familiar with, eight they had never seen before. They had no feedback, and a limit of 5 min per problem. The error under consideration is incorrect diagnosis, including failure to diagnose.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 0.34    | Simulator         |
| Petro-chemical [11]          | Before the lifeboat is lowered to sea level, one of the internal checks that the coxswain must perform involves checking the air support system. This check was omitted. If the system is not checked and the crew find out after they have abandoned, there is the possibility of ingress of smoke. Lack of air will cause asphyxiation.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 0.037   | Simulator         |
| Calculator errors [17]       | An experiment is performed on a group of engineers selected from a large industry in America to investigate their performance in using electronic calculators. Each engineer is given a test and is required to calculate answers to numerical problems. The test consists of 10 problems ranging from very simple ones to more complex ones. Each problem is selected from the manual supplied with the calculator. The error of concern is that the engineer will make an error in the block of 10 problems.                                                                                                                                                                                                                                                                                                                                  | 0.27    | Experimental data |
| Aviation [18]                | A new system for data transfer and display for use by air traffic controllers is being evaluated in a simulation study. The simulation is based on normal flight operations at Heathrow Airport. The new systems displays information to the controller on a single cathode ray tube which features a touch sensitive overlay on the display through which data is entered and controlled. The controller performs his normal duties, using the new system to transmit, obtain and update information. There are three possible errors the controller can commit: miss touches, if the controller misses the area which should be touched; error touches, when the controller touches the wrong area; and illegal touches, when the controller touches a label out of sequence. In this scenario the air traffic controller makes a miss error. | 0.064   | Simulator data    |
| Nuclear [19]                 | A locally operated valve has a rising stem and position indicator. An auxiliary operator, while using written procedures to check a valve line-up, fails to realise that the valve is not in the right position, after a maintenance person has performed a procedure intended to restore it to its proper position.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 0.003   | Expert judgement  |

As a general rule, for making an initial screening of HEPs, any tasks which are complex or time-pressured may be considered as a first approximation to be in the 0.1–1.0 probability region. Routine tasks may be considered to be in the 0.01–0.1 region, and well-designed procedural tasks with good training, human interface and checking can be considered to be in the 0.001–0.01 region. Such values should not themselves be used in an actual HRA, but are ‘rules of thumb’ applied by assessors when first considering a range of tasks that may require quantification.

#### *Human Performance–Limiting Values*

Dependence has already been introduced via the THERP approach, which considers dependence between subsequent tasks. However, it is possible that a risk assessment failure path contains several errors, all of which must happen, for example human error 1 (e.g. an initiating event error) and human error 2 (a failure by an operator to detect the event) and human error 3 (a failure to diagnose the event) and human error 4 (a failure to recover the event). Even using dependence modelling as used in THERP, it is sometimes the case that such sequences (called ‘cutsets’ in risk analysis terms) end up with a combined HEP that is unrealistically low (e.g. 0.000001). Incident and accident experience shows us in fact that although we try hard to model all failures, there are often surprises in store for us – things happen which were not predicted (e.g. an unpredicted error of commission) or for reasons that were assumed to be implausible. This is also the case in hardware reliability and risk assessment, and in that domain cutsets are prevented from being optimistic by using limiting values or other methods (e.g. beta factors). The same approach is necessary in HRA, and so human performance–limiting values (HPLVs) were developed (Kirwan et al., 1990). Effectively, the general rule is not to allow the human error component of a cutset to go below 0.00001 or  $10^{-5}$  unless robust human error data exist to substantiate such low values or unless there are extremely compelling qualitative arguments. This also corresponds with recent HRA guidance in the US Nuclear Power industry (USNRC, op. cit.) and one of the second-generation HRA techniques discussed shortly (MERMOS).

#### *Validity of the HRA Quantification Approaches*

Towards the end of the 1980s, there were a number of techniques (around a dozen) in use in nuclear, process and chemical and offshore oil and gas sectors of industry, and in an effort to clarify the situation, a series of validations were carried out (Kirwan, 1997b,c). In such exercises, HRA assessors were given task descriptions and asked to use their techniques to predict results for which the validators had data (and such data were not known by the assessors). Where prediction accuracy fell below a reasonable proportion (e.g. most HEPs needed to be predicted within a factor of 10, and preferably a factor of 3), the technique was invalidated. In such exercises, for example, THERP, APJ and HEART received positive validations, whereas SLIM’s predictive accuracy depended critically on certain factors (notably ‘calibration’ of the expert judgements) and was less assured. One approach, in particular, which predicted human performance effectively solely as a function of time available to carry out the task (called time–response curve) was invalidated (Dolby, 1990; Kantowitz and Fujita, 1990).

To show how validation works, the results of a large-scale UK validation (Kirwan, 1999) are summarised here, which focused on the three quantification techniques THERP, HEART and JHEDI. Thirty UK HRA practitioners took part in the exercise, and each assessor independently used only one of the three techniques (very few assessors were practitioners in more than one technique). Each assessor quantified HEPs for 30 tasks, and for each task there was the following information:

- General description of the scenario
- Inclusion of relevant PSF information in the description
- Provision of simple linear task analysis
- Provision of diagrams where necessary and relevant
- Statement of exact human error requiring quantification

Each assessor had 2 days to carry out the assessments, and experimental controls were exercised, so that the assessors were working effectively under invigilated examination conditions. For each of the 30 tasks, the HEP was known to the experimenter, but not to any of the assessors. Tasks were chosen to be relevant to nuclear power and reprocessing industries, since all of the assessors were currently working in these two areas. A large proportion of the data were from real recorded incidents, with the data range spanning five orders of magnitude (i.e. from 1.0 to  $1\text{E}-5$ ).

The analysis of all the data (i.e. all 895 estimated HEPs – there were five missing values) showed a significant correlation between estimates and their corresponding true values (Kendall's coefficient of concordance:  $Z = 11.807$ ,  $p < 0.01$ ). This supported the validity of the HRA quantification approach as a whole, especially as no assessors or outliers were excluded from these results. The analysis of all data for individual techniques shows a significant correlation in each case (using Kendall's coefficient of concordance): THERP  $Z = 6.86$ ; HEART  $Z = 6.29$ ; JHEDI  $Z = 8.14$ , all significant at  $p < 0.01$ .

Individual correlations for all subjects are shown in Table 30.13. There are 23 significant correlations (some significant at the  $p < 0.01$  level) out of a possible 30 correlations. This is a very positive result, again supporting the validity of the HRA quantification approach.

Table 30.14 shows that there is an overall average of 72% precision (estimates within a factor of 10) for all assessors, irrespective of whether they were significantly correlated or not. This figure includes all data estimates, even the apparent outliers that have been identified in the study. This is therefore a reasonably good result, supporting HRA quantification as a whole. Furthermore, no single assessor dropped below 60% precision in the study. The precision within a factor of 3 is approximately 38% for all techniques. This is a fairly high percentage given the required precision level of a factor of 3. The degree of optimism and pessimism

**TABLE 30.13**  
**Correlations for Each Subject for the Three Techniques**

|    | <b>THERP</b> | <b>HEART</b> | <b>JHEDI</b> |
|----|--------------|--------------|--------------|
| 1  | 0.615**      | 0.577**      | 0.633**      |
| 2  | 0.581**      | 0.558**      | 0.551**      |
| 3  | 0.540**      | 0.473**      | 0.533**      |
| 4  | 0.521**      | 0.440**      | 0.452**      |
| 5  | 0.437**      | 0.389*       | 0.436**      |
| 6  | 0.311*       | 0.370*       | 0.423*       |
| 7  | 0.298        | 0.351*       | 0.418*       |
| 8  | 0.297        | 0.347*       | 0.401*       |
| 9  | 0.254        | 0.217        | 0.386*       |
| 10 | 0.078        | 0.124        | 0.275        |

\*\*Significant  $p < 0.01$ ; \*significant  $p < 0.05$ .

**TABLE 30.14**  
**Precision for the Three Techniques for All HRA Assessors**

|       | <b>Factor 3 (%)</b> | <b>Factor 10 (%)</b> | <b>Optimistic (%)</b> | <b>Pessimistic (%)</b> |
|-------|---------------------|----------------------|-----------------------|------------------------|
| THERP | 38.33               | 72.00                | 13.67                 | 13.33                  |
| HEART | 32.67               | 70.33                | 11.00                 | 18.33                  |
| JHEDI | 43.67               | 75.00                | 3.33                  | 21.33                  |



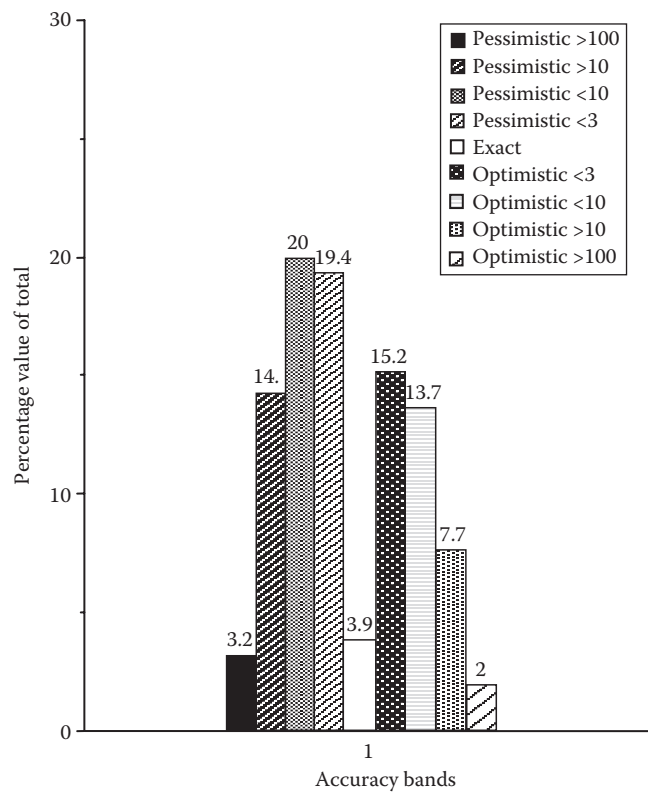


FIGURE 30.6 Overall HRA optimism and pessimism.

is not too large, as also shown in the histogram in Figure 30.6, with only a small percentage of estimates at the extreme optimistic and pessimistic ends of the histogram (i.e. greater than a factor of a hundred from the actual estimate). Certainly there is room for improvement, but the optimism and pessimism are not in themselves dominating the results, and estimates were more likely to be pessimistic (17.5% of the total number of estimates) than optimistic (9.7% of the total number of estimates), which means that assessors erred on the side of caution. The highest and lowest precision values for the techniques within a factor of 3 and within a factor of 10 are shown in Table 30.15.

The overall results were therefore positive, with a significant overall correlation of all estimates with the known true values, with 23 individual (assessor) significant correlations, and a general acceptable precision range of 60%–87%, the average precision being 72%. These results lend support for the HRA quantification part of the HRA process in general.

TABLE 30.15  
Precision for the Three Techniques

|       | Highest Precision |                  | Lowest Precision |                  |
|-------|-------------------|------------------|------------------|------------------|
|       | Factor of 3 (%)   | Factor of 10 (%) | Factor of 3 (%)  | Factor of 10 (%) |
| THERP | 56.67             | 80.00            | 10.00            | 63.33            |
| HEART | 46.67             | 76.67            | 16.67            | 60.00            |
| JHEDI | 66.67             | 86.67            | 26.67            | 70.00            |

### *Second-Generation HRA*

In the early 1990s, a prominent HRA expert (Dougherty, 1990) suggested, based on the experiences of the accidents mentioned earlier, that most HRA approaches did not pay enough attention to context, that is, to the detailed scenarios people found themselves in. Essentially, the argument was that considering human reliability in an abstracted format of fault and event trees was insufficient to capture the local situational factors that would actually dictate human behaviour and lead to success or failure. This occurred at the same time as a growing concern in the United States in particular with EOCs, wherein a human in the system does something that is erroneous and not required by procedures (e.g. shutting off emergency feed water to the reactor in TMI, disconnecting safety systems while running reactivity experiments in Chernobyl). Such errors of intention, relating to a misconception about the situation, are, as already noted earlier, severely hazardous to industries such as nuclear power and aviation. Incident experience in the United States was suggesting that it was these types of rare errors, whose very unpredictability made them difficult to defend against, which were of most real concern.

Therefore, work on a set of so-called 'second-generation HRA' techniques began in the early to mid-1990s. The most notable of these were a technique for human error analysis (ATHEANA: Cooper et al., 1996) and the cognitive reliability error analysis method (CREAM: Hollnagel, 1998) and MERMOS (Le Bot, 2003). For ATHEANA, the qualitative part of the method is intensive and allows for deep analysis of context, and is used for identification of safety critical human interventions. It has recently been adapted to look at fire scenarios in nuclear power plants. CREAM is a more straightforward quantification technique that has had some applications in the nuclear power domain. MERMOS is used in France and focuses on the conditions that need to coincide to give rise to errors of commission.

The term second generation really refers to two aspects: the first is a more profound analysis of contextual factors (including interacting PSF) to ensure the accuracy and validity of the resultant HEPs, and the second is the ability to predict errors of commission (ATHEANA) or the conditions that lead to them (MERMOS).

Though the methods mentioned earlier have all been applied already in various safety assessments, there is still more work to be done regarding the evidence of empirical validation compared to 'first-generation' techniques (like HEART or THERP). The level of application and work to be done can also be seen from a special issue of a key reliability journal on HRA and EOCs (Straeter, 2004).

A recent large-scale international benchmark exercise (Bye et al., 2011), occurring in two steps, initially suggested that second-generation techniques were better at modelling what operators would actually do, while first-generation techniques like THERP and HEART were more accurate on predicting the HEPs. However, although the results of the second step are not yet publicly available, it is understood that in the second step of the validation the second-generation techniques' quantitative performance was improved.

The second-generation HRA approach is clearly still evolving. Taken to one extreme, there is less focus on individual errors, and more focus on determining what factors can combine to lead to an intrinsically unsafe situation, wherein errors of intention become increasingly likely. This concept has some resonance proponents of the emergent field of 'resilience engineering' (Hollnagel et al., 2006).

This does not mean that second-generation HRA research should stop, nor does it mean that HRA is useless without this function of predicting errors of commission – HRA can still identify and predict most types of errors, and occasionally errors of commission too, especially via human HAZOP techniques for error identification, and if need be, formal expert judgement techniques for quantification or else HRA techniques such as MERMOS and ATHEANA. What it does mean for system developers and risk managers is that other methods and measures must be utilised, adopting a defence-in-depth strategy for managing human error. This includes maintaining a good level of human factors adequacy in the operator interfaces, environment and support systems (training, procedures, work

organisation, etc.), as well as a good safety culture that will avoid or reduce violations and help detect the seeds of errors of commission long before they manifest themselves in accidents.

In practical terms, first-generation techniques are still useful. Rather than second-generation techniques being seen as an attempt to replace first-generation HRA tools, a more likely future relationship is one of complementarity. For example, SPAR-H or coupled with ATHEANA could be an effective partnership for the US nuclear industry, and in the United Kingdom a similar partnership could operate between NARA and MERMOS. Such partnerships may offer a more flexible and resource efficient approach to HRA, and a safer world in which to live.

### *Safety Culture and Human Reliability*

*Safety culture* concerns the degree to which the organisational culture of the installation's organisation puts safety as the over-riding priority. It came into being as an area of study and improvement following the 1986 Chernobyl disaster. If everyone in the organisation is personally concerned with safety, mindful of safety and willing to challenge practices if they believe them to be unsafe, then clearly this has an impact on error reduction, error recovery enhancement and violation avoidance, as well as watching out for errors of commission.

Violations in particular are inextricably bound together with the concept of safety culture. An installation with a good safety culture should not exhibit rule violations, whereas one with a poor safety culture will exhibit routine and even perhaps, under certain circumstances, extreme violations. The concern with rule violations is similar to that for mistakes – violations can lead to failure of multiple safety systems and barriers. As an example, a risk assessment may assume that a fully programmed machine with multiple interlocks will not be used in 'maintenance mode' (where the safety interlocks are disabled) for normal operations. Such a 'routine' violation would not only violate the procedures of the installation, but would seriously violate the assumptions of the risk analysis, such that its predictions of failure for such a system could easily be incorrect by one or more orders of magnitude.

When carrying out HRA and risk assessment, there are usually general assumptions made about the level of training and adherence to procedures. If there is a poor safety culture, the working assumptions of the risk assessment, at a certain point, will become invalid. This is to an extent what happened with both Bhopal and Chernobyl. No risk assessor would have assumed that so many safety systems would have been switched off or disabled.

There is a link therefore between HRA (and risk assessment) and safety culture – ideally safety culture should be measured periodically (e.g. every 3–4 years) in an organisation to verify that the 'fundamentals' of adequate behaviour and safe attitudes at work, which many would consider simply professional conduct, are being maintained.

More recent accidents such as the offshore oil and gas Deepwater Horizon accident, as well as the nuclear power disaster at Fukushima Daichi in Japan, continue to emphasise the importance of safety culture. Safety culture survey approaches can sometimes unearth such issues, particularly if using interview or workshop techniques as well as questionnaires where people from the local operating culture and management are asked confidentially whether there are any significant risks that are not being addressed sufficiently. This approach can find so-called accidents waiting to happen. In this sense safety culture acts as a collective and predictive version of the critical incident technique discussed earlier.

## **HRA, ETHICS, JUST CULTURE AND THE RESILIENCE PERSPECTIVE ON HUMAN ERROR**

An important standpoint of HRA, as mentioned at the outset, is that it does not allocate blame. Almost always, accidents and incidents and human errors are the result of system design and operational constraints, complexities and pressures. Most errors are unintentional, and in the majority

of tragic accidents, for example in aviation or rail, the human operators at the sharp end are not alive to defend themselves after the event. HRA therefore, as with risk assessment, aims to be dispassionate – the search is always for errors and their consequences, their causes, their relative importance and effective (and sustainable) remedies. Even when an intentional error occurs or is predicted (e.g. a violation of procedures), the assessor is looking for deeper causes in the organisational culture (e.g. poor safety culture). Only when malice is intended, and this is exceptionally rare in work-related systems, would the real connotation of blame come into the equation, and at such a point this would be out of the hands of HRA and pass into those of criminal investigation. This standpoint of HRA practitioners is more critical than for risk analysts, since the former will be interviewing operators and management from the system under investigation (whether an existing system or one that is on the drawing board). HRA analysts are predicting how a workforce, or a pilot, or a train driver, or a surgeon, could make serious errors, possibly leading to deaths – this naturally has to be handled sensitively. HRA analysts must convey (and believe) that they are there to support these operators by finding system vulnerabilities and helping them and the parent organisation overcome them.

This point is practical as well as ethical. If people think they will be prosecuted for honest mistakes, then they will not report them. This means that the organisation and the wider community will not be able to learn from such errors or even discover them until a large-scale accident occurs. This is usually far too late to learn, both in terms of the losses that have already occurred (e.g. fatalities and loss of equipment, etc.) and the financial and public confidence losses that will follow for the organisation and perhaps even the industry concerned in the wake of large-scale accidents. HRA therefore aims to work within what is known as a ‘just culture’ framework (Dekker, 2007).

Recent works in the resilience area (e.g. Hollnagel et al., 2013) once again highlight the systemic view of accidents and human performance in complex systems. While resilience appears most useful today in the understanding of systemic accidents, and appears today to be practically more useful for incident investigation rather than proactive and predictive risk assessment and management, this area is definitely ‘one to watch’ for HRA and PSA analysts. Advocates of resilience would argue for a paradigm shift towards a new way of looking at and assessing risks, though this author does not believe its predictive capability is ready yet, and also does not see why it needs to be resilience ‘or’ PSA/HRA, since these two safety camps are not mutually exclusive.

Human reliability, particularly in mature industrial systems where the right balance between humans and engineered safety functions has been found, is often exceptionally good. HRA can therefore also be seen as an approach to continuous improvement, ensuring that the human operators (and their managers) are able to be both safe and effective – the two usually go hand-in-hand. As a closing point, therefore, as resilience often argues, it has to be said that most of the time, people ‘get it right’, and accidents are rare, and generally systems are very safe *and* productive, and would not achieve this without the human operator, and we must also focus on why things go right, not only why they fail.

## CONCLUSION

HRA started in the 1960s and has grown to be a useful tool in many industries where the human operator plays a critical safety-related role. It is now used qualitatively and/or quantitatively in nuclear, chemical, offshore oil and gas, defence, aviation and air traffic, space and medical sectors. This is a testament to the success of the overall approach.

HRA has been used in many assessments to decide whether the design is safe or reliable enough and to determine how to improve the human reliability in the system’s performance. Although HRA’s evolution is far from over, there is a useful methodology and toolkit available for any industry wishing to focus on supporting, protecting and enhancing its major asset – the human in the system.

## REFERENCES

- Bye, A., Lois, E., Dang, V., Parry, G., Forester, J., Massaiu, S., Boring, R. et al. (2011) International HRA empirical study – Phase 2 report results from comparing HRA method predictions to simulator data from SGTR scenarios, Vol. 2, USNRC Report IA/2016. Washington, DC: U.S. Nuclear Regulatory Commission.
- Cooper, S.E., Ramey-Smith, A.M., Wreathall, J., Parry, G.W., Bley, D.C., Luckas, W.J., Taylor, J.H. and Barriere, M.T. (1996) A technique for human error analysis (ATHEANA) – Technical basis and method description, NUREG/CR-6350. Washington, DC: U.S. Nuclear Regulatory Commission.
- Cox, S.J. and Tait, N.R.S. (1991) *Reliability, Safety and Risk Management*. Oxford, U.K.: Butterworth-Heinemann.
- Dekker, S. (2007) *Just Culture: Balancing Safety and Accountability*. Aldershot, U.K.: Ashgate.
- Dolby, A.J. (1990) A comparison of operator response times predicted by the HCR model with those obtained from simulators. *International Journal of Quality and Reliability Management*, 7(5), 19–26.
- Dougherty, E.M. (1990) Human reliability analysis – Where shouldst thou turn? *Reliability Engineering and System Safety*, 29, 283–299.
- Embrey, D.E. (1986) SHERPA – A systematic human error reduction and prediction approach. Paper presented at the *International Topical Meeting on Advances in Human Factors in Nuclear Power Systems*, Knoxville, TN.
- Embrey, D.E., Humphreys, P.C., Rosa, E.A., Kirwan, B. and Rea, K. (1984) SLIM-MAUD: An approach to assessing human error probabilities using structured expert judgement, NUREG/CR-3518. Washington, DC: U.S. Nuclear Regulatory Commission.
- Embrey, D.E., Kontogiannis, T. and Green, M. (1994) *Preventing Human Error in Process Safety*. New York: Centre for Chemical Process Safety (CCPS), American Institute of Chemical Engineers.
- Gertman, D.I. and Blackman, H. (1994) *Human Reliability and Safety Analysis Data Handbook*. Chichester, U.K.: John Wiley & Sons.
- Gertman, D.I., Blackman, H.S., Marble, J., Byers, J., Haney L.N. and Smith, C. (2005) *The SPAR-H Human Reliability Analysis Method*, NUREG/CR-6883. Washington, DC: U.S. Nuclear Regulatory Commission.
- Gibson, W.H., Dennis, C., Thompson, K., Mills, A. and Kirwan, B. (2012) Tailoring the HEART technique for application in the rail industry. In *Advances in Safety, Reliability and Risk Management (ESREL 2011)*, Berenguer, C., Grall, A. and Guedes Soares, C. (Eds.). London, U.K.: CRC Press.
- Greenstreet Berman. (2001) *Preventing the Propagation of Error and Misplaced Reliance on Faulty Systems: A Guide to Human Error Dependency*, Offshore Technology Report 2001/053. London, U.K.: UK Health & Safety Executive, HMSO.
- Hollnagel, E. (1998) *CREAM: Cognitive Reliability Error Analysis Method*. Oxford, U.K.: Elsevier Science Ltd.
- Hollnagel, E., Braithwaite, J. and Wears, R.L. (2013) *Resilient Health Care*. Aldershot, U.K.: Ashgate.
- Hollnagel, E., Woods, D.D. and Leveson, N. (2006) *Resilience Engineering: Concepts and Precepts*. Aldershot, U.K.: Ashgate.
- Kantowitz, B. and Fujita, Y. (1990) Cognitive theory, identifiability, and human reliability analysis. *Reliability Engineering and System Safety*, 29(3), 317–328.
- Kirwan, B. (1992) Human error identification in HRA. Part 2: Detailed comparison of techniques. *Applied Ergonomics*, 23(6), 371–381.
- Kirwan, B. (1994) *A Guide to Practical Human Reliability Assessment*. London, U.K.: Taylor & Francis Group.
- Kirwan, B. (1997a) The development of a nuclear chemical plant human reliability management approach. *Reliability Engineering and System Safety*, 56, 107–133.
- Kirwan, B. (1997b) Validation of human reliability assessment techniques: Part 1 – Validation issues. *Safety Science*, 27(1), 25–41.
- Kirwan, B. (1997c) Validation of human reliability assessment techniques: Part 2 – Validation results. *Safety Science*, 27(1), 43–75.
- Kirwan, B. (1998) Human error identification techniques for risk assessment of high risk systems – Part 1: Review and evaluation of techniques. *Applied Ergonomics*, 29(3), 157–177.
- Kirwan, B. (1999) Some developments in human reliability assessment. In *The Occupational Ergonomics Handbook*, Karwowski, W. and Marras, W.S. (Eds.). London, U.K.: CRC Press, pp. 643–666.
- Kirwan, B. (2006) Safety informing design. *Safety Science*, 45, 155–197.
- Kirwan, B. and Ainsworth, L.A. (Eds.) (1992) *A Guide to Task Analysis*. London, U.K.: Taylor & Francis Group.
- Kirwan, B., Basra, G. and Taylor-Adams, S. (June 1997) TA: A computerised human error database for human reliability support. In *IEEE Conference on Human Factors and Power Plants*, Orlando, FL.

- Kirwan, B. and Gibson, W.H. (2007) CARA: A human reliability assessment tool for air traffic management – Technical basis and preliminary architecture. In *The Safety of Systems*, Redmill, F. and Anderson, T. (Eds.). London, U.K.: Springer-Verlag, pp. 163–178.
- Kirwan, B., Gibson, H., Kennedy, R., Edmunds, J., Cooksley, G. and Umbers, I. (2004) Nuclear action reliability assessment (NARA): A data-based HRA tool. In *Probabilistic Safety Assessment and Management*, Spitzer, C., Schmocker, U. and Dang, V.N. (Eds.). London, U.K.: Springer, pp. 1206–1211.
- Kirwan, B. and Kennedy, R. (11–12 June 2001) Assessing safety and usability of Air Traffic Management (ATM) systems using a HAZOP approach. In *Human Error & Safety System Development (HESSD 2001)*, Linköping, Sweden.
- Kirwan, B., Kilner, A., Gibson, H., Piccione, D. and Sawyer, M. (2011) The development and application of CARA – A HRA tool for Air Traffic Management systems. In *Advances in Safety, Reliability and Risk Management (ESREL 2011)*, Berenguer, C., Grall, A. and Guedes Soares, C. (Eds.). London, U.K.: CRC Press.
- Kirwan, B., Martin, B.R., Rycraft, H. and Smith, A. (1990) Human error data collection and data generation. *International Journal of Quality and Reliability Management*, 7(4), 34–66.
- Le Bot, P. (2003) Methodological validation of MERMOS by 160 analyses. In *Proceedings of the International Workshop Building the New HRA: Errors of Commission from Research to Application*, NEA/CNSI/R(2002)3. Issy-les-Moulineaux, France: OECD.
- Reason, J.T. (1990) *Human Error*. Cambridge, U.K.: Cambridge University Press.
- Reason, J.T. (1998) *Managing the Risks of Organisational Accidents*. Aldershot, U.K.: Ashgate.
- Seaver, D.A. and Stillwell, W.G. (1983) *Procedures for Using Expert Judgement to Estimate Human Error Probabilities in Nuclear Power Plant Operations*, NUREG/CR-2743. Washington, DC: U.S. Nuclear Regulatory Commission.
- Shorrock, S.T. and Kirwan, B. (2002) Development and application of a human error identification tool for air traffic control. *Applied Ergonomics*, 33, 319–336.
- Sträter, O. (Ed.) (2004) Human reliability analysis: Data issues and errors of commission. Special Issue of *Reliability Engineering and System Safety*, 83(2), 127–128.
- Swain, A.D. (1989) *Comparative Evaluation of Methods for Human Reliability Analysis*, GRS-81. Garching, Germany: Gesellschaft für Reaktorsicherheit (GRS)mbH.
- Swain, A.D. and Guttmann, H.E. (1983) *A Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*, Nureg/CR-1278. Washington, DC: USNRC.
- Topmiller, D.A., Eckel, J.S. and Kozinsky, E.J. (1984) *Human Reliability Databank for Nuclear Power Plant Operations*, Report NUREG/CR-2744. Washington, DC: USNRC.
- USNRC. (2005) *Good Practices for Implementing Human Reliability Analysis*, NUREG-1792. Washington, DC: U.S. Nuclear Regulatory Commission.
- Williams, J.C. (1988) A data-based method for assessing and reducing human error to improve operational performance. In *IEEE Conference on Human Factors in Power Plants*, Monterey, CA, 5–9 June 1988, pp. 436–450.



---

# 31 Incident Reporting and Analysis

*Brendan Ryan*

## CONTENTS

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| Introduction.....                                                       | 821 |
| Definitions and Terminology .....                                       | 822 |
| Reasons for Reporting and Investigation of Incidents .....              | 823 |
| Sources of Information on Events.....                                   | 824 |
| Need for Structure in the Collection and Analysis of Incident Data..... | 825 |
| Outputs and Solutions .....                                             | 829 |
| Conclusions: Final Thoughts .....                                       | 830 |
| Appendix.....                                                           | 831 |
| References.....                                                         | 833 |

## INTRODUCTION

Things go wrong. It could be a relatively isolated event involving a single person, such as a fork lift truck overturning in a warehouse. Or it could be a much larger event, made up of many individual but potentially interacting events and circumstances, such as can be found in any one of a number of high-profile accidents in nuclear, process control or transport domains. These types of incidents can receive much coverage in the academic literature, published reports and the media or popular press (e.g. the nuclear incident at Chernobyl in 1986; the explosion at Buncefield in 2005; the train crash at Ladbroke Grove in 1999).

It seems very appropriate for both smaller and larger events to try to find out what went wrong; to ensure that people learn from experiences (of themselves and others), to stop things going wrong again. Investigators and managers often try to do this, with varying degrees of success. This lack of success is possibly related to the complex nature of some of these types of events and the fact that some things go wrong because of the seemingly random, often interacting effects of different events and circumstances that are part of the overall accident. Problems also arise from weaknesses in the design and application of systems of reporting, investigation, analysis and learning from feedback. For example, Cooke and Rohleder (2006), commenting on the ‘Challenger’ space shuttle accident, referred to problems where organisations did not

synthesise and share the information from separate ‘precursor’ incidents with the relevant people across the organisation so that appropriate action could be taken to reduce the risk of disaster.

How should relevant information be collected? How should that information be used?

This chapter provides a practical focus for the ergonomist, safety specialist or manager, with signposts and relevant links to the different aspects of theory that are relevant in the process of learning from events. This review tries not to be influenced or constrained by any particular theory or perspective. However, there is certainly an ergonomics/human factors (E/HF) focus, with an understanding of the performance of people in the system and a focus on identifying ‘mismatches between system demands and operator behaviour’ (Rasmussen, 1987). This type of E/HF approach to investigation and analysis reflects the question posed by Dekker – ‘Why did their decision make sense to them at the time?’ (Stoop and Dekker, 2012). It goes beyond searching for departures from operating instructions (e.g. in some highly structured tasks) or attempts to determine the causes of accidents or attribute blame (Brown, 1995). Instead, this approach recognises the potential value of diversity in performance of different operators and considers error as ‘the effect of human variability in an unfriendly environment’ (e.g. Kjellén, 1984a,b; Rasmussen, 1987) (see Chapter 30 of this book for more information on human error theory). Therefore, investigations should seek to produce more error-tolerant systems. This can be achieved by broadening the investigation process beyond the investigation of what people did, to include analysis of the system and consider alternatives of what people could have done to meet task and system goals without incident or with the potential for recovery from error (Leplat, 1987; Stoop and Dekker, 2012).

The chapter begins with a brief introduction to some terminology and relevant definitions, outlining reasons for identifying and reporting on different types of events. Common sources of data on accidents and incidents are presented, along with explanation of how these data can be used. This includes the analysis of findings from investigation of individual events and how these findings can be aggregated in accident/incident reporting systems. The chapter also includes some introduction to the types of analysis and classification that are needed, based on theories of causation and accident models. However, this review is not exhaustive. Enough material is provided to justify the collection of broad-ranging information, so that those involved in investigation and analysis maintain an openness to new findings about their system. Finally, examples of outputs from the analysis are discussed. These include the investigation report and the content and use of information databases, both of which are integral to the sharing of lessons from the investigation of incidents. Throughout, there is an emphasis on outlining how good, reliable data can be collected, stored and used within organisations. Some of the pitfalls, biases and challenges in learning from events are discussed.

By way of an extended example, this chapter concludes with details of how to ‘design, implement, maintain and operate a system for near miss reporting and analysis’, drawing from the work of Van der Schaaf and Wright (2005, third edition of this volume).

## DEFINITIONS AND TERMINOLOGY

It is difficult to establish unambiguous definitions of relevant terms such as accident or incident. These terms are often used differently, either within the literature or within industry. Earlier reviews have referred to definitions of accident in common usage (in dictionary definitions, with connotations of chance or luck, the association with injury or damage to property, or the sudden, unplanned or unforeseen nature of different types of events) (Brown 1995; Hollnagel, 2005). Hovden et al. (2010) refer to an accident as ‘a hazard materialising in a sudden, probabilistic event (or chain of events) with adverse consequences (injuries)’.

Hollnagel (2005) uses the magnitude and severity of the outcome to distinguish between accident and incident, defining an accident as ‘an unforeseen event or occurrence, which results in serious property damage or injury, possibly even loss of life’, whereas an incident may result in lesser consequences (e.g. minor injury or property damage). This type of continuum can be extended to include other associated terms such as dangerous occurrences (e.g. as used in UK RIDDOR; HSE, 2013), as prescribed events that have not resulted in a reportable injury, though clearly had the potential to do so and unsafe acts (actions which do not comply with principles of safe behaviour) (Hollnagel, 2005).

There are various examples within the literature which attempt to illustrate relative frequencies of these events of different magnitude (e.g. pyramid/iceberg models based on early work by Heinrich et al., 1980). However, the lack of data on near misses (events which might have become incidents and accidents in different circumstances) and differences in definitions used in different industries (Brazier, 1994) leads to questions about the reported ratios between the numbers of different types of events (Hollnagel, 2005). The higher frequencies of less severe events and the relative ease of collecting more complete and less biased reports from those involved in events without serious injury or property damage (Van der Schaaf and Wright, 2005) can offer opportunities for learning.

For the purpose of this chapter, it is clear that there are different types of events, with different magnitudes, affecting people and property in different ways. There are likely to be opportunities for learning from events, both small and large in nature. Organisations need to understand the characteristics of the events that are particular to the organisation, to provide these learning opportunities.

## REASONS FOR REPORTING AND INVESTIGATION OF INCIDENTS

There are various reasons for reporting and investigating an incident. Reports and investigations may be influenced by legal requirements, particularly in relation to more serious events. Examples of these types of circumstances are available in the UK RIDDOR legislation (HSE, 2013) and associated guidance, including requirements to report events with major injuries (e.g. fractures) or where an event leads to an absence from work for more than seven days. Whilst there is no legally binding requirement for notification and investigation of some other (generally less serious) types of events, the opportunities for organisational learning arise from developing a better understanding of the circumstances surrounding single events (e.g. new types of incidents) or understanding trends in aggregated statistics. Nevertheless, care should be taken when attempting to draw conclusions from data on the frequencies of incidents. For example, in circumstances in which the numbers of incidents are found to remain fairly consistent over time (Dien et al., 2012a; Franzten, 2004), there can be a number of explanations which are linked to different aspects of organisational performance (e.g. continued high numbers of incidents and risk to workers or evidence of a reporting culture and effective reporting within the organisation). Furthermore, effort in analysing and learning from large amounts of information can be high (Berentsen and Holmboe, 2004). Organisations should therefore clarify what types of events should be reported and how, based on priorities that have been established for learning within the organisation. Practical aspects of establishing a good reporting system are dealt with later in this chapter.

Fundamental to the whole process of reporting is the way in which people identify and respond to incidents or events in the workplace. 'Events' can differ in terms of how conspicuous they are, both in the nature of the events or their consequences, from the perspective of the people involved and any observers. When an event occurs, there needs to be recognition on the part of several people that this should be reported. Webb et al. (1989) explain how a system of selectively permeable filters, influenced by personal and organisational factors, determines how many accidents or injuries will be reported. These include the worker, who may or may not label the event as a work-related accident or may be reluctant to report this for a number of reasons. Furthermore, supervisors, medical staff and safety personnel make decisions as to whether to refer the matter to others. These decisions are potentially influenced by organisational processes and incentive programmes (Pransky et al., 1999).

Having identified an accident, investigations are carried out from different perspectives. Those with an interest can include regulators (safety enforcement or police), industry personnel (engineers, managers, employees, unions) and researchers (from safety, psychological, sociological, engineering, E/HF, organisational disciplines) (Dien et al., 2012a; Saleh et al., 2010). Each can have different agendas (e.g. considering legal responsibilities, allocating blame, determining facts, looking for explanations, identifying preventative measures or developing theory) (LeCoze, 2008; Roed-Larsen

and Stoop, 2012), with the potential to pick up issues relating to power, the normalisation of deviations and the impacts of complexity in organisations (Dien et al., 2012a; Vaughan, 1996). These different perspectives can influence the approach to the investigation (e.g. the independence of the investigator can be a positive factor, but the merits of this can potentially be traded off against their knowledge of the environment) (Roed-Larsen and Stoop, 2012). Furthermore, the methods that are used can influence the depth of data collection and analysis. Outputs from investigation and analysis can be used in different ways. For example, Katsakiori et al. (2009) refer to the different interests of investigators (who investigate and analyse) and statisticians (who compile and register data in databases and official statistics).

## SOURCES OF INFORMATION ON EVENTS

The scale of the incident can influence the effort that is applied in collecting relevant information on what happened and in processing data to enable organisational learning. Investigations may be carried out by an individual, a team or a much larger commission (e.g. CAIB, 2003) in some major accidents. Outputs from investigations may include a detailed accident report (which responds to the terms of reference or a remit for an investigation or inquiry) or the types of summary data that are found commonly in incident databases. Whatever the scale of any investigation, many of the principles are the same. Interviews or written statements (from those involved or those witnessing an incident) may be one of the primary ways of collecting information on historical events. Ryan et al. (2009) have reviewed literature and conducted analysis on the factors affecting the content of different types of self-reports in this type of post-accident investigation situation in the railway environment. These factors can include the person's awareness during an event, their memory for the event and the type of rationalisation that may occur, difficulties reflecting on and articulating experiences and the impact of the types of questions used by investigators. There is clearly important information that can be collected from those involved in accidents and much of this is not available from other means. However, careful consideration needs to be given to the types of prompts or questions that are used in eliciting this information.

Similarly, Drury et al. (2002) found that when events were reported, only about a third of relevant facts were identified and a small proportion of these facts about the events appeared in summary reports to management. Improvements could be made by careful design of reporting forms so that the information that is relevant to the purpose of the investigation (e.g. to the E/HF specialist if informing future system or work design) is collected routinely. Some types of information may be easy to collect (e.g. biographical or demographic data), but may be of limited value if it does not help to explain why an event occurred (Brown, 1995).

Sanne (2008) encouraged the use of 'storytelling', which is likely to occur naturally between colleagues as a way of collecting more details about real-life incidents that occurred on the railway. Sanne suggested that this might help to combat problems of under-reporting of incidents, but acknowledged that ways of focusing on the systemic causes of incidents would be needed as these are not typically reported within stories. Detailed collection and analysis of text and descriptions within stories can complement other methods and be part of a multi-method approach, particularly from a research perspective.

Relevant information is also available in various forms of archival data. These sources provide opportunities to learn from records related to real events and the circumstances in which these occur, with minimal interference to the working system (Bizantz and Drury, 2005). These archival sources can include outputs from previous investigations (e.g. accident reports or other injury reports), formal reporting records (e.g. RIDDOR) or anonymous reports of incidents (e.g. confidential incident reporting for safety (CIRAS), in GB rail). Business-related documents that are compiled for other purposes (e.g. plans and records of locations, equipment, annual reports, sales figures) can be used

for understanding the size, structure and geography of an organisation, and help in the development of a reporting system (Bizantz and Drury, 2005).

Data from reports and investigations of incidents are often stored in databases and used subsequently for the basis of analysis (e.g. comparison by different locations or across jobs, tasks or sectors in a population). Examples include industry information systems (e.g. the safety management information system [SMIS] in the GB railway system) or national/international systems (e.g. Swedish Occupational Injury Information System [Andersson and Lagerlof, 1983; Kemmlert and Lundholm, 2001], major accident reporting system [MARS] in the European Union, the Swedish Traffic Accident Data Acquisition [STRADA] and the Aircraft Accident Statistics and Knowledge [AASK] in the United States, the United Kingdom and Australia [Lindberg et al., 2010]).

Databases of incidents are valuable, but the contents are likely to be an incomplete picture of the evidence and local circumstances. Not all incidents will be known and recorded (especially where there is no 'visible' injury). There may be limited facts, especially in summary reports and database entries. There is likely to be more detailed knowledge of a smaller number of the more serious events, which may have been selected for in-depth investigation and analysis. It is important to be aware of the restrictions that can be imposed by classification systems for incident reporting, as well as the subjective biases and errors when reporting and recording the findings from reports or investigations of the incidents (Bizantz and Drury, 2005).

## **NEED FOR STRUCTURE IN THE COLLECTION AND ANALYSIS OF INCIDENT DATA**

Outputs from investigations are commonly recorded within individual reports of events. These can be text based, describing in detail how an accident occurred. This type of descriptive detail can be an essential part of the investigation process and important for the learning of the organisation (organisational learning will be discussed later in greater detail). Detailed, descriptive reports can be largely inaccessible to many within organisations due to the length and amount of detail in the documents. Developing and using an underlying structure can help with standardising investigation processes and communication of relevant data and outputs (e.g. numbers of incidents in different locations, trends over time), both within and across organisations (Roed-Larsen and Stoop, 2012). Okstad et al. (2012) explain how standardised approaches for the collection of descriptive details can save time and resources, lead to more credible results and compensate for any weaknesses in competence in the investigation teams.

From a practical point of view, it is necessary to consider the reliability of classification or coding of data from the investigation process. Classification of relevant characteristics of incidents is often carried out by more than one person (e.g. in industrial settings when coding incidents in a company database or in research activities where there are several investigators). Olsen (2013) has pointed out that the measurement of reliability in coding is considered in too few studies or, where it is carried out, uses unsuitable measures of reliability (see Chapter 1 for further discussion of the concept of reliability in more detail). Ross et al. (2004) recommend demonstrating 'inter-rater consensus', using an index of concordance (number of agreements/number of agreements plus number of disagreements), rather than using other measures of 'consistency'. A fuller discussion of the issues involved in coding is available in the literature (Kozlowski and Hattrup, 1992; Wallace et al., 2002), including some useful, practical examples (Olsen, 2013; Olsen and Shorrock, 2010).

Accident theories, models and methods can be considered as a hierarchy. Relevant theory or safety concepts influence the development of accident models and these models in turn influence the selection of accident analysis methods. In choosing suitable analysis methods, it is essential to give consideration to the context and circumstances of work operations.

There have been various reviews of the history and development of accident theory and safety concepts that underpin the process of analysis. For example, Saleh et al. (2010) refer to change from the emphasis on front-line staff to greater focus on the organisational origin of causes. Hovden et al. (2010) review Perrow's normal accident theory (with impacts of complexity and tight coupling) and the notion of high-reliability organisations (with continuous training, use of redundancy, heightened awareness of the unexpected, resilience and ability to absorb, preserve and recover, and decision-making at appropriate system levels [LaPorte and Consolini, 1991]). The deviation concept is outlined by Kjellén (1984a) and Kjellén and Larsson (1981), including deviations from norms in different phases of an accident (initiator phases – with a lack of control, concluding phase – with a loss of control, leading to an injury phase). Stoop and Dekker (2012) explain the changing focus in the area of accident modelling, risk management concepts (Rasmussen, 1997) and Hollnagel's (2006) resilience engineering concept, with a greater focus more recently on systems analysis, rather than event analysis. Dien et al. (2012b) outline how investigation has evolved, from focus on technical aspects (1970s), human error (1980s), interactions in socio-technical systems (1990s, 2000s), up to the present day, where there is a focus on the relationships between organisations.

Accident models or simplified representations influence the way we look for information on incidents, how we understand and analyse this information, identify interrelations between factors and conditions, and produce recommendations and communicate findings (Hovden et al., 2010). No single model can be used to explain all accidents and models need to be selected on the basis of the context of the work situation and the objectives or purpose of the analysis (Saari, 1986). There are many reviews of different accident models in existence (e.g. Andersson, 1991; Katsakiori et al., 2009; Khanzode et al., 2012; Sklet, 2004), classifying models according to different theoretical or practical groupings (e.g. pure chance, domino, process, energy, human information processing, epidemiological, system and macro-ergonomic-based models). These models can consider the whole system, processes within a system or look at a system from different perspectives (e.g. taking a human information processing approach; Corlett and Gilbank, 1978; Hale and Hale, 1970). Several of the models distinguish between events and contributing factors, or other synonymous terms such as determining factors or conditions (Benner, 1975; Kjellén and Larsson, 1981; Tuominen and Saari, 1982; Williamson and Feyer, 1990). Events have been described as being active, specific to incidents and often in sequences, whereas determining factors or conditions describe states or circumstances (Sklet, 2004) (e.g. environmental conditions, work processes, procedures) and can be more stable components in the system (Andersson, 1991; Tuominen and Saari, 1982; Väyrynen et al., 1994). As a result of the relative stability of determining factors (in comparison with events), there have been arguments for focusing preventative measures on the control of the determining factors.

Katsakiori et al. (2009) emphasise the links between the accident models and methods, explaining how it is important to select an accident investigation model that fits a set of circumstances and then determine the methods that are appropriate for use with the selected model. A range of different investigation methods were evaluated, to align these, where possible, with underlying theoretical classifications of accident models (e.g. sequential, human information processing and systemic) (Katsakiori et al., 2009). Requirements of an accident analysis method were also presented, based on earlier work by Wagenaar and van der Schrier (1997). Sklet (2004), in a detailed and highly cited review, placed particular emphasis on the characteristics of a number of models and methods for organising evidence in the analysis of accidents. A list of requirements for accident investigation and analysis is given in Table 31.1.

Several researchers have commented on the different competencies that may be needed within an investigation or investigation teams (e.g. the need for domain specific knowledge, technological knowledge or systematic investigation experience) (Hovden et al., 2011). It is also necessary for investigators to keep up to date with changing philosophies, models and methods of investigation



**TABLE 31.1****Requirements for Methods for Accident Investigation and Analysis**

An accident analysis method should

- Have a clear scope for analysis (e.g. whether it should focus at the level of the work and technological system, or more broadly at influences from government and regulators)
- Be influenced by a model or group of models
- Provide a detailed description of the accident, including a visual representation of the accident sequence if appropriate
- Search for and reveal underlying causes
- Contribute to understanding of prevention (e.g. safety barriers)
- Help in generating recommendations
- Give consideration to practical aspects, such as the level of education and training that is needed to use the method
- Be valid and reliable

*Source:* Based upon Katsakiori et al. (2009), Sklet (2004) and Wagenaar and van der Schrier (1997).

and be able to respond to challenges imposed by new technologies and complex systems (Dien et al., 2012b; Okstad et al., 2012).

Researchers have made distinctions between the types of failures and influencing factors in different types of events (Menckel, 1990) and questioned whether different theories and models are needed to explain events in different contexts. More complex accident models, analysis systems and safety indicators can be required in high hazard domains, whereas simpler energy or barrier-based models (e.g. Gibson [1961] and Haddon's [1968] 10 strategies for loss prevention) can be appropriate for more straightforward, general occupational accidents (Dien et al., 2012b; Hovden et al., 2010). Carroll and Fahlbruch (2011) recommended the need for simple and usable tools for managers and investigators. Jørgensen (2011) focused on use of a practical tool for safety officers or managers to learn from incidents in circumstances that are more mundane (i.e. events which are often considered as 'nothing special' or out of the ordinary).

Table 31.2 contains some examples of analysis tools and methods, based on reviews from a number of researchers (Dien et al., 2012a; Herrera and Woltjer, 2010; Katsakiori et al., 2009; Khanzode et al., 2012; Okstad et al., 2012; Sklet, 2004).

Different methods of investigation and analysis focus on different aspects of the system. Some analysis tools attempt to characterise aspects of human information processing, decision-making and human error (e.g. TRACER, Shorrock and Kirwan, 2002; HFACS, Shappell and Wiegmann, 2000), using taxonomies to help with the classification of relevant data from the investigation (see Chapter 30 of this book for a review of some human error identification techniques, including TRACER). These methods can help to broaden the investigation to consider questions such as: Was there a problem with the presented information? Did people expect different information? Was the wrong rule chosen?

The ACCIMAP method focuses more on the representation of evidence in a diagrammatic format, including the potential influences across different levels within the sociotechnical system (Svedung and Rasmussen, 2002). Emphasis is placed on factors where there is potential for improvement, rather than the collection of all facts about the incident. This method reflects the need to ensure there is a suitable control structure for risk management. Variations of the ACCIMAP are possible (e.g. aggregating a number of different scenarios, showing relationships between different organisations, showing information flow among decision-makers).

System Theoretic Accident Model and Processes (STAMP) (Leveson, 2004) is an example of a system-based method for analysis. The accident is viewed as a dynamic and adaptive process, with actors within different levels of a system and with interactions across different levels of the system, also with potential for influences beyond the organisation (LeCoze, 2008; Leveson, 2004). Accidents occur as the result of inadequate control of safety constraints (including control actions and feedback), which

**TABLE 31.2****Examples of Methods and Tools That Help to Provide Structure in the Collection and Analysis of Accident-Related Data****Method and Brief Description**

*The accident evolution and barrier function (AEB) method* – based on work of Svensson, with particular focus on interactions between human and technical systems at all parts of an accident sequence.

*ACCIMAP* – based on the work of Svedung and Rasmussen, with emphasis on a number of different levels of decision-making that are important in risk management.

*Barrier analysis* – barriers that were in place and whether they performed or not.

*Change analysis* – examining planned and unplanned changes that resulted in the undesirable outcomes.

*Cognitive Reliability Error Analysis Method (CREAM)* – from Hollnagel, including causes, effects and the contexts in which incidents take place.

*Events and causal factors charting and analysis* – including events, conditions and causal factors.

*Event tree analysis* – analysing the sequences that follow an initiating event.

*Fault tree analysis* – a graphic approach to determining the causes of an event. Aspects of fault trees (causal chains) and event trees (consequences) have been combined in bow-tie methods by Bellamy and colleagues.

*Functional Resonance Analysis Method (FRAM)* – from Hollnagel, an analytic method considering dynamic interactions (including non-linear dependencies, variability in functions involving people, technology and organisations).

*HFACS* – by Wiegmann and Shappell, building on Reason's concept of active and latent conditions/failures, with taxonomies to support the identification of failures/conditions at four levels (acts, preconditions, supervision and organisational influences).

*Influence diagrams* – determining basis events, the decision and actions that influenced these, and the relationship of these decisions and actions to organisational factors.

*Management Oversight Risk Tree (MORT)* – based on work by Johnson, using an analytical tree/graphical checklist to identify problems in control factors and management system factors.

*Man, Technology and Organisation (MTO)* – based on work of Røllenhagen and Bento, using a combination of event and causal diagrams, change analysis and barrier analysis and associated checklists.

*Multilinear events sequencing* – from Benner, identifying a series of events that can lead to an accident.

*Root cause analysis* – a systematic process to identify underlying problems in the safety management system.

*Sequential Timed Events Plotting (STEP)* – based on Benner's multilinear sequences of events, with a worksheet and matrix to plot events for different actors along a horizontal timescale.

*STAMP* – from Leveson, system – based method, where the accident is seen as a dynamic and adaptive process, with actors and interactions across different levels of the system.

*STEP* – from Hendrick and Benner, describing an accident as a process of sequentially timed events.

*Systematic Cause Analysis Technique (SCAT)* – a chart and checklist, based on Bird's Loss Causation model, to ensure that different aspects of the accident are considered, including factors contributing to lack of control, basic causes, immediate causes, the incident and the types of loss.

*TRACEr* – from Shorrock and Kirwan, including error classification using taxonomies to support the identification of external and internal error modes, the cognitive framework, psychological error mechanisms and error detection and recovery.

*TRIPOD* – based on the work of Groeneweg and Reason, with a strong emphasis on organisational causes of incidents, which led to the TRIPOD Beta tool, merging a hazard and effects model (HEMP) with the different components and assumptions in the TRIPOD model (containing latent failures, preconditions, active failures, failed controls or defences, which lead to an accident).

maintain the conditions that the system was designed for, through control loops between different levels of the system (Okstad et al., 2012; Saleh et al., 2010). There is therefore a focus on the conditions that allow events to occur (Leveson, 2011), rather than the events themselves. LeCoze (2008) notes that a typical limitation of accident analysis is that event-based investigations may not collect the necessary information for this type of analysis.

Examples of application of a number of these investigation and analysis methods can be found in the literature. Baysari et al. (2008, 2009) apply HFACS and TRACER in railway contexts. ACCIMAPs have been applied within several domains and contexts (Jenkins et al., 2010; Vicente and Christoffersen, 2006; Waterson, 2009). Ouyang et al. (2010) use STAMP in relation to a railway accident. Harris and Li (2011) use classifications from HFACS (of unsafe acts, pre-conditions for unsafe acts, unsafe supervision and organisational influences and ‘categories of control flaws’ from STAMP in the investigation of an aircraft mid-air collision). Herrera and Woltjer (2010) demonstrate the analytical nature of FRAM in a study of an aircraft incident to illustrate ‘why’ dynamic interactions occur within the sociotechnical systems, in comparison with the descriptive nature of ‘what happened’ using STEP.

Structure is clearly important for the analysis of accidents and for the future use of data arising from investigations. This can be essential for harmonising investigation methods within and across different companies or sectors (Okstad et al., 2012). However, it is necessary to achieve the right balance between structured, analytic approaches (how and why things occurred) and more descriptive approaches (what happened) (Herrera and Wojalter, 2010; LeCoze, 2008). In some circumstances a structured approach (i.e. normative – where data that are collected should fit a structure or model), could put some constraints on the process of collecting and analysing the data. It may be more appropriate in other situations to apply a descriptive approach (where models are derived from the collection of detailed, descriptive data) (LeCoze, 2008).

Some common weaknesses of investigation and analysis have been identified by Dien et al. (2012a). These weaknesses of methods include not going far enough back in time, difficulties dealing with complexity in systems, inability to incorporate enough of the qualitative detail, limited organisational/inter-organisational and sociological coverage, such that there is insufficient understanding of how organisations are working, their weaknesses and resilience within everyday working.

The increasing complexity in work systems (e.g. linked to ICT, remote control of processes) needs to be considered when choosing appropriate models and methods for investigation (Okstad et al., 2012). The changing nature of work systems is considered in greater detail by Hovden et al. (2010) and earlier by Wilpert (2009) and Mayhew et al. (1997), considering how factors such as globalisation, immigration, technological developments, automation, deregulation and sub-contracting could impact on safety, the changing nature of accidents and new types of hazards.

Rasmussen (1997) originally posed the question of whether there are adequate models ‘in the present dynamic society’. Some of these difficulties (e.g. dynamic, complex systems) mean that attempting to learn from the past is not enough (Leveson, 2011). Some of the more recent accident investigation methods (e.g. STAMP, FRAM) start to tackle these concerns, with more focus on system constraints than linear causes and greater understanding of the organisational context that allows an accident to happen (LeCoze, 2008). However, these newer methods still need to be evaluated (e.g. for their effectiveness and ease of use by investigators) (Herrera and Wojalter, 2010).

## OUTPUTS AND SOLUTIONS

One of the primary purposes of investigation and incident reporting systems is to enable organisational learning (see Chapter 30 of this book for discussion of industry blind spots and the need for feedback from safety events). Organisational learning can occur at different levels in an organisation (e.g. at an individual level in the event of specific incidents, or a team level where colleagues can learn from the incidents that are experienced by others; Hollnagel, 2005). Furthermore, learning can occur as part of a programme which includes the analysis of trends within aggregated data.

Too often learning does not happen, especially if there is no culture of learning in the organisation, inadequate feedback about events, or too much focus on allocating responsibility for incidents (Hovden et al., 2011). Carroll and Fahlbruch (2011) pose a number of relevant questions when commenting on weaknesses in how organisations learn from incidents. Do analysis methods not discover underlying causes? Do organisations not learn from experience? Is learning not occurring in the right places?

There are a number of weaknesses that are implied within these questions, which need to be considered by organisations. Organisations need to focus their attention on different aspects of the investigation process, resulting in prompt completion of a high-quality initial investigation report. The organisation needs to be sensitive to the opportunities for learning from within (e.g. the selection of relevant events for analysis) and external to the organisation (e.g. inter-organisational learning, including better communications with other organisations such as unions, suppliers, contractors, regulators and legislators). This will enable understanding of potential weaknesses in shared mental models of the system, and sharing of new knowledge on common problems or the use of new technologies (Carroll and Fahlbruch, 2011). It is also necessary to consider how outputs from investigations are disseminated and how any resulting interventions are supported to ensure that they are effective (Hollnagel, 2005).

Much of the effort in incident investigation and safety improvement usually focuses on attempts to identify the right causes, with few studies covering the difficulties in solving the problems that have been found (Lundberg et al., 2010). Recommendations for safety improvement can be limited to efforts to 'fix what is fixable' (in the opinion of the investigators). Lundberg et al. (2010) explain that this can be influenced by a number of factors, including the expertise of the investigators; the resources given to the investigation; the availability of data; political considerations (within or external to the organisation); stopping rules for analysis; existing plans for fixes that are in progress, what others decide to fix; what is thought to be possible, relevant and known to work in different contexts and domains; what is cost-beneficial.

Kirwan (2011) provides an interesting example of failure in learning in a situation where system-related changes that were implemented in response to the identification of trends in incidents in air traffic control were not necessarily the right interventions. This example shows how problems can recur in a different (potentially more dangerous) way, in this case because of pressure on capacity in the air traffic system. This is illustrative of the challenges of specifying, implementing and monitoring the effectiveness of countermeasures in increasingly complex systems, with the potential for new manifestations of risk.

## CONCLUSIONS: FINAL THOUGHTS

Things go wrong and will continue to go wrong. This will likely occur in new and different ways in the future. There is value in conducting investigations and using the findings (from investigations and other incident data), to strengthen the safety management system in the organisation. There are various examples of theory and lots of accident models and investigation methods to support the analysis. It is necessary to understand the context in which the accident occurs, for the correct selection and application of these investigation and analysis tools.

A detailed example of components of an incident reporting/analysis system is given in the Appendix, based on the content from Van der Schaaf and Wright (2005, third edition of this volume). This example clarifies three goals of a reporting system and the steps that are needed in designing such a system. Preconditions for implementing the system in an organisation (e.g. the type of support needed) are outlined, along with some opportunities for use and dissemination of the outputs. The components from this example can be considered and applied more widely in the design and development of other forms of accident and incident reporting and analysis systems.

## APPENDIX\*

### GOALS OF A NEAR-MISS REPORTING SYSTEM

A near-miss reporting system has three goals:

- Modelling and learning about new failure modes – to improve understanding, identify error-inducing factors and strengthen recovery promoting factors. However, there can be limitations in the use of pre-coded taxonomies/classifications of errors.
- Monitoring the frequency of known failures (and factors influencing these, combinations of factors influencing these) for trends.
- Maintaining alertness and motivating staff – for use in training and safety promotion (e.g. for recurring problems or to be able to identify and respond to new combinations of circumstances).

### DESIGNING REPORTING SYSTEMS

This will include various decisions which will need to be taken by the company about how to administer the system (e.g. in-house or by a third party) and will depend on factors such as the company culture, willingness to report, need for confidentiality. Van der Schaaf (1992) provides a framework which outlines the modules that are needed in a near-miss reporting system, and how these contribute to the goals outlined earlier. A summary of the modules is included next:

*Detection*, including recognition and reporting: What types of events should be recorded will need to be based on how the company see the importance of one or more of the three goals. This will include encouragement to staff to report, raising awareness of the system (e.g. how it operates and the practicalities of making it easy to report), how reports should be made (e.g. whether these are anonymous, by telephone, structured forms or free-text descriptions, electronic forms or notification in person) and what benefits can be gained from reporting.

*Selection* of events for analysis: Determining which reports to analyse in depth (dependent partly on reporting rates), which will include minimal analysis and which will be considered for monitoring purposes (e.g. repeated incidents). This will be based upon a clear rationale for any decisions that are taken (e.g. taking account of the consequences or focusing on different types of events, such as maintenance incidents or new events).

*Description* of all relevant hardware, human and organisational factors: Including analysis on an individual basis (e.g. modelling). This stage may also need a structure to enable comparison across multiple events (monitoring).

*Classification* according to a socio-technical model: Also carried out on an individual basis (if modelling is the goal) or using a taxonomy. This can enable comparison with other incidents and statistical analysis of a database, to determine the most common factors and combinations of factors. Minimum and optional requirements of a taxonomy for this purpose are given in Wright (2002) and Shorrock (2002).

*Computation*, statistical analysis to uncover patterns: With root causes which have been identified by classification, entered into a database for further analysis of the factors contributing to multiple events.

*Interpretation and implementation* – Translating statistical results into corrective and preventative measures, including support and suggestions for how factors influencing error can be eliminated and how recovery opportunities can be strengthened.

---

\* Developed from Van der Schaaf and Wright (2005) – third edition of this volume.

*Evaluation* – Measuring the effectiveness of proposed measures after implementation, to ensure that the system achieves its goals. However, the effectiveness of interventions can be hard to monitor due to the influence of other changes in the industry (other interventions, changes in culture and rules) and time lags between interventions and effects. It is important that steps are taken to build trust in the system, so that those who need to report into the system see the value of this. It is also necessary for management to be willing to make improvements and support the implementation of measures that are based on the information that is received from the reporting system.

### PRECONDITIONS FOR IMPLEMENTING A SUCCESSFUL REPORTING SYSTEM

*Management support:* Management must understand the goals of the system and company structures and cultures need to be aligned with the system. Successful schemes may need to rely on anonymity (this is reliant on collecting good quality information first time because of no opportunities for follow-up) and confidentiality (where additional re-assurance can remove fears of ridicule or embarrassment),

*No-blame reporting:* Managers would need to be able to show forgiveness to achieve this (where punitive actions will not result, and with the intention of finding out what is actually happening, without dwelling on who is breaking the rules). Immunity from discipline cannot be guaranteed and is not necessary as an integral part of a system. Reason (2001) refers to the desire for transition to a ‘just’ culture (needing to decide where to draw the line between events that warrant discipline and those that do not, considering motivations of people as well as actions), rather than a no-blame culture. A ‘reporting culture’ may be a step towards a just culture (which may be difficult to achieve in practice).

*Training and supporting staff:* Staff will need support in a number of areas, to develop awareness and trust in the system (e.g. awareness of the aims of the system and how the system operates, how reports are processed, what happens to the information that is provided, how results will be fed back). This can be achieved through various media, such as letters, presentations, videos, briefings and demonstrations. It may be necessary to provide incentives for reporting (though this will need to be done carefully to minimise biases in the data, avoid trivial reports and a significant drop off in reports after incentives are removed). Some barriers to reporting include (Van der Schaaf and Kanse, 2004) fear (of discipline or ridicule), acceptance of risk (this is part of the job), uselessness of reporting (management don’t take any notice anyway) and practicalities (such as the time it takes for reporting). The success of a scheme can be dependent upon a combination of the design of the system, how it is implemented and the culture into which it is being placed.

*Defining feedback routes to staff:* This can include acknowledging receipt of a report and provision of information on what will happen next (e.g. whether someone can expect to be contacted to provide more information). Feedback of results of any individual analysis of the event or other relevant analysis of aggregated data could be by direct contact or through a regular publication.

### DIFFERENT LEVELS OF ORGANISATIONAL LEARNING

Decisions need to be taken about the level at which near-miss reporting systems need to operate, as part of the process of design and implementation (e.g. operation at departmental, company, regional and national levels will have implications for issues of compatibility between hardware and software systems, as well as the degree of control of the type of data that can be collected and shared between different organisations or parts of an organisation). Larger national systems may have advantages of more data and greater opportunities for learning, but smaller more local systems may have greater levels of trust and more specific prevention measures.



## REFERENCES

- Andersson, R. 1991. *The Role of Accidentology in Occupational Injury Research*, The National Institute of Occupational Health, Solna, Sweden.
- Andersson, R. and Lagerlof, E. 1983. Accident data in the new Swedish information system on occupational injuries, *Ergonomics*, 26(1), 33–42.
- Baysari, M.T., Caponecchia, C., McIntosh, A.S. and Wilson, J.R. 2009. Classification of errors contributing to rail incidents and accidents: A comparison of two human error identification techniques, *Safety Science*, 47, 948–957.
- Baysari, M.T., McIntosh, A.S. and Wilson, J.R. 2008. Understanding the human factors contribution to railway accidents and incidents in Australia, *Accident Analysis and Prevention*, 40, 1750–1757.
- Benner, L. 1975. Accident investigations: Multilinear events sequencing methods, *Journal of Safety Research*, 7(2), 67–73.
- Berentsen, R. and Holmboe, R.H. 2004. Incidents/accidents/classification and reporting in Statoil, *Journal of Hazardous Materials*, 111, 155–159.
- Bizantz, A.M. and Drury, C.G. 2005. Applications of archival and observational data, in J.R. Wilson and E.N. Corlett (eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology*, 3rd edn., Taylor & Francis Group, London, U.K., pp. 61–82.
- Brazier, A.J. 1994. A summary of incident reporting in the process industry, *Journal of Loss Prevention in the Process Industries*, 7, 243–248.
- Brown, I. 1995. Accident reporting and analysis, in J.R. Wilson and E.N. Corlett (eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology*, 2nd edn., Taylor & Francis Group, London, U.K., pp. 969–992.
- Buncefield. 2014. <http://www.buncefieldinvestigation.gov.uk/> (Accessed on April 17, 2014).
- CAIB. 2003. Columbia accident investigation board report, NASA, Washington, DC, retrieved from [http://www.nasa.gov/columbia/home/CAIB\\_Vol1.html](http://www.nasa.gov/columbia/home/CAIB_Vol1.html) (Accessed April 17, 2014).
- Carroll, J.S. and Fahlbruch, B. 2011. The gift of failure: New approaches to analysing and learning from events and near-misses. Honouring the contributions of Bernhard Wilpert, *Safety Science*, 49, 1–4.
- Chernobyl. 2014. <http://news.bbc.co.uk/1/shared/spl/hi/guides/456900/456957/html/nn1page1.stm> (Accessed on April 17, 2014).
- Cooke, D.L. and Rohleder, T.R. 2006. Learning from incidents: From normal accidents to high reliability, *System Dynamics Review*, 22, 213–239.
- Corlett, E.N. and Gilbank, G. 1978. A systemic technique for accident analysis, *Journal of Occupational Accidents*, 2, 25–38.
- Dien, Y., Dechy, N. and Guillaume, E. 2012a. Accident investigation: From searching direct causes to finding in-depth causes – Problem of analysis or/and of analyst? *Safety Science*, 50, 1398–1407.
- Dien, Y., Dechy, N. and Stoop, J. 2012b. Perspectives regarding industrial events investigation, *Safety Science*, 50, 1377–1379.
- Drury, C.G., Ma, J. and Woodcock, K. 2002. *Measuring the Effectiveness of Error Investigation and Human Factors Training*, FAA Office of Aviation Medicine, Washington, DC. [http://www.faa.gov/about/initiatives/maintenance\\_hf/library/documents/media/human\\_factors\\_maintenance/measuring\\_the\\_effectiveness\\_of\\_error\\_investigation\\_and\\_human\\_factors\\_training.pdf](http://www.faa.gov/about/initiatives/maintenance_hf/library/documents/media/human_factors_maintenance/measuring_the_effectiveness_of_error_investigation_and_human_factors_training.pdf). (Accessed April 17, 2014).
- Franzén, C. 2004. Tango on Asymptote, in *SRA-E Annual Conference*, Paris, France, 15–17 November 2004.
- Gibson, J.J. 1961. The contribution of experimental psychology to the formulation of the problem of safety – A brief for basic research, in: *Behavioral Approaches to Accident Research*, Association for the Aid of Crippled Children, New York, pp. 77–89.
- Haddon, W. 1968. The changing approach to epidemiology, prevention and amelioration of trauma, *American Journal of Public Health*, 58(8), 1431–1438.
- Hale, A.R. and Hale, M. 1970. Accidents in perspective, *Occupational Psychology*, 44, 115–121.
- Harris, D. and Li, W.-C. 2011. An extension of the human factors analysis and classification system for use in open systems, *Theoretical Issues in Ergonomics Science*, 12(2), 108–128.
- Health and Safety Executive. 2013. RIDDOR – Reporting of injuries, diseases and dangerous occurrences regulations, Statutory Instruments 2013, No. 1471, The Stationery Office Ltd., Norwich, U.K.
- Heinrich, H.W., Petersen, D. and Roos, N. 1980. *Industrial Accident Prevention: A Safety Management Approach*, 5th edn., McGraw-Hill Book Company, New York.
- Herrera, I.A. and Woltjer, R. 2010. Comparing multi-linear (STEP) and systemic (FRAM) method for accident analysis, *Reliability Engineering and System Safety*, 95, 1269–1275.

- Hollnagel, E. 2005. Learning from failures: A joint cognitive systems perspective, in J.R. Wilson and E.N. Corlett (eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology*, 3rd edn., Taylor & Francis Group, London, U.K., pp. 901–916.
- Hollnagel, E., Woods, D.D. and Leveson, N. 2006. *Resilience Engineering: Concepts and Precepts*, Ashgate, Aldershot, U.K.
- Hovden, J., Albrechtsen, E. and Herrera, I.A. 2010. Is there a need for new theories, models and approaches to occupational accident prevention, *Safety Science*, 48, 950–956.
- Hovden, J., Størseth, F. and Tinmannsvik, R.K. 2011. Multilevel learning from accidents – Case studies in transport, *Safety Science*, 49, 98–105.
- Jenkins, D.P., Salmon, P.M., Stanton, N.A. and Walker, G.H. 2010. A systemic approach to accident analysis: A case study of the Stockwell shooting, *Ergonomics*, 53(1), 1–17.
- Jørgensen, K. 2011. A tool for safety officers investigating “simple” accidents, *Safety Science*, 49, 32–38.
- Katsakiori, P., Sakellariopoulos, G. and Manatakis, E. 2009. Towards an evaluation of accident methods in terms of their alignment with accident causation models, *Safety Science*, 47, 1007–1015.
- Kemmert, K. and Lundholm, L. 2001. Slips trips and falls in different work groups – With reference to age and from a preventative perspective, *Applied Ergonomics*, 32, 149–153.
- Khanzode, V.V., Maiti, J. and Ray, P.K. 2012. Occupational injury and accident research: A comprehensive review, *Safety Science*, 50, 1355–1367.
- Kirwan, B. 2011. Incident reduction and risk migration, *Safety Science*, 49(1), 11–20.
- Kjellén, U. 1984a. The deviation concept in occupational accident control – I. Definition and classification, *Accident Analysis and Prevention*, 16(4), 289–306.
- Kjellén, U. 1984b. The role of deviations in accident causation and control, *Journal of Occupational Accidents*, 6, 117–126.
- Kjellén, U. and Larsson, T.J. 1981. Investigating accidents and reducing risks – A dynamic approach, *Journal of Occupational Accidents*, 3, 129–140.
- Kozlowski, S.W.J. and Hattrup, K. 1992. A disagreement about within-group agreement: Disentangling issues of consistency versus consensus, *Journal of Applied Psychology*, 77, 161–167.
- LaPorte, T.R. and Consolini, P.M. 1991. Working in practice but not in theory: Theoretical challenges of high-reliability organizations, *Journal of Public Administration Research and Theory*, 1, 19–47.
- LeCoze, J.-C. 2008. Disasters and organisations: From lessons learnt to theorising, *Safety Science*, 46, 132–149.
- Leplat, J. 1987. Accidents and injury production: Methods of analysis, in J. Rasmussen, K. Duncan and J. Leplat (eds.), *New Technology and Human Error*, John Wiley, Chichester, U.K., pp. 133–142.
- Leveson, N. 2004. A new model for engineering safer systems, *Safety Science*, 42, 237–270.
- Leveson, N. 2011. Applying systems thinking to analyze and learn from events, *Safety Science*, 49, 55–64.
- Lindberg, A.-K., Hansson, S.O. and Rollenhagen, C. 2010. Learning from accidents – What more do we need to know? *Safety Science*, 48, 714–721.
- Lundberg, J., Rollenhagen, C. and Hollnagel, E. 2010. What you find is not always what you fix – How other aspects than causes of accidents decide recommendations for remedial actions, *Accident Analysis and Prevention*, 42, 2132–2139.
- Mayhew, C., Quinlan, M. and Ferris, R. 1997. The effects of subcontracting/outsourcing on occupational health and safety: Survey evidence from four Australian industries, *Safety Science*, 25(1–3), 163–178.
- Menckel, E. 1990. *Intervention and Cooperation. Occupational Health Services and Prevention of Occupational Injuries in Sweden*, National Institute of Occupational Health, Solna, Sweden.
- Okstad, E., Jerson, E. and Tinmannsvik, R.K. 2012. Accident investigation in the Norwegian petroleum industry – Common features and future challenges, *Safety Science*, 50, 1408–1414.
- Olsen, N.S. 2013. Reliability studies of incident coding systems in high hazard industries: A narrative review of study methodology, *Applied Ergonomics*, 44(2), 175–184.
- Olsen, N.S. and Shorrock, S.T. 2010. Evaluation of the HFACS-ADF safety classification system: Inter-coder consensus and intra-coder consistency, *Accident Analysis and Prevention*, 42, 437–444.
- Ouyang, M., Hong, L., Yu, M.-H. and Fei, Q. 2010. STAMP-based analysis on the railway accident and accident spreading: Taking the China-Jiaoji railway accident for example, *Safety Science*, 48, 544–555.
- Perrow, C. 2007. *The Next Catastrophe: Reducing Our Vulnerabilities to Natural, Industrial, and Terrorist Disasters*, Princeton University Press, Princeton, NJ.
- Pransky, G., Snyder, T., Dembe, A. and Himmelstein, J. 1999. Under-reporting of work related disorders in the workplace: A case study and review of the literature, *Ergonomics*, 42(1), 171–182.
- Rasmussen, J. 1987. The definition of human error and taxonomy for technical system design, in J. Rasmussen, K. Duncan and J. Leplat (eds.), *New Technology and Human Error*, John Wiley, Chichester, U.K., pp. 23–30.

- Rasmussen, J. 1997. Risk management in a dynamic society: A modelling problem, *Safety Science*, 27, 183–213.
- Reason, J.T. 2001. *Human Error Reduction Operation, for Railway Safety Limited*, The Vision Consultancy, London, U.K.
- Roed-Larsen, S. and Stoop, J. 2012. Modern accident investigation – Four major challenges, *Safety Science*, 50, 1392–1397.
- Ross, A.J., Wallace, B. and Davies, J.B. 2004. Technical note: Measurement issues in taxonomic reliability, *Safety Science*, 42, 771–778.
- Ryan, B., Hutchings, J. and Lowe, E. 2009. An analysis of the content of questions and responses in incident investigations: Self reports in the investigation of signals passed at danger (SPADs), *Safety Science*, 48, 372–381.
- Saari, J. 1986. Accident epidemiology, in M. Karvonen and M.I. Mikheev (eds.), *Epidemiology of Occupational Health*, European Series No. 20, WHO Regional Publications, Geneva, Switzerland, pp. 299–316.
- Saleh, J.H., Marais, K.B., Bakolas, E. and Cowlagi, R.V. 2010. Highlights from the literature on accident causation and system safety: Review of major ideas, recent contributions, and challenges, *Reliability Engineering and System Safety*, 95, 1105–1116.
- Sanne, J. 2008. Incident reporting or storytelling? Competing schemes in a safety-critical and hazardous work setting, *Safety Science*, 46, 1205–1222.
- Shappell, S.A. and Wiegmann, D.A. 2000. The human factors analysis and classification system – HFACS, DOT/FAA/AM-00/7, U.S. Department of Transportation Federal Aviation Administration, Washington, DC.
- Shorrock, S.T. 2002. Error classification for safety management – Finding the right approach, in Presented at *Investigation and Reporting of Incidents and Accidents*, University of Glasgow, Glasgow, 17–20 July 2002.
- Shorrock, S.T. and Kirwan, B. 2002. Development and application of a human error identification tool for air traffic control, *Applied Ergonomics*, 33, 319–336.
- Sklet, S. 2004. Comparison of some selected methods for accident investigation, *Journal of Hazardous Materials*, 111, 29–37.
- Stanton, N.A. and Walker, G.H. 2011. Exploring the psychological factors involved in the Ladbroke Grove rail accident, *Accident Analysis and Prevention*, 43(3), 1117–1127.
- Stoop, J. and Dekker, S. 2012. Are safety investigations pro-active? *Safety Science*, 50(6), 1422–1430.
- Svedung, I. and Rasmussen, J. 2002. Graphic representation of accident scenarios: Mapping system structure and the causation of accidents, *Safety Science*, 40, 397–417.
- Tuominen, R. and Saari, J. 1982. A model for analysis of accidents and its application, *Journal of Occupational Accidents*, 4, 263–273.
- Van der Schaaf, T.W. 1992. Near miss reporting in the chemical processing industry, PhD thesis, Technische Universiteit Eindhoven, Eindhoven, the Netherlands.
- Van der Schaaf, T.W. and Kanse, L. 2004. Biases in incident reporting databases: An empirical study in the chemical process industry, *Safety Science*, 42, 57–67.
- Van der Schaaf, T.W. and Wright, L. 2005. Systems for near miss reporting and analysis, in J.R. Wilson and E.N. Corlett (eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology*, 3rd edn., Taylor & Francis, London, U.K., pp. 877–899.
- Vaughan, D. 1996. The challenger launch decision: Risky technology, culture and deviance at NASA, University of Chicago Press, Chicago, IL.
- Väyrynen, S., Pekkarinen, A. and Tornbeg, V. 1994. Some links between accidents, postural load and accessibility in chemical plant maintenance, *Safety Science*, 18, 125–133.
- Vicente, K.J. and Christoffersen, K. 2006. The Walkerton *E. coli* outbreak: A test of Rasmussen's framework for risk management in a dynamic society, *Theoretical Issues in Ergonomics Science*, 7(2), 93–112.
- Wagenaar, W.A. and van der Schrier, J. 1997. Accident analysis – The goal, and how to get there, *Safety Science*, 26, 25–33.
- Wallace, B., Ross, A., Davies, J., Wright, L. and White, M. 2002. The creation of a new minor event coding system, *Cognition, Technology and Work*, 4, 1–8.
- Waterson, P. 2009. A systems ergonomics analysis of the Maidstone and Tunbridge Wells infection outbreaks, *Ergonomics*, 52(10), 1196–1205.
- Webb, G.R., Redman, S., Wilkinson, C. and Sanson-Fisher, R.W. 1989. Filtering effects in reporting work injuries, *Accident Analysis and Prevention*, 21(2), 115–123.
- Williamson, A. and Feyer, A.-M. 1990. Behavioural epidemiology as a tool for accident research, *Journal of Occupational Accidents*, 12, 207–222.
- Wilpert, B. 2009. Impact of globalization on human work, *Safety Science*, 47, 727–732.
- Wright, L. 2002. The analysis of UK railway accidents and incidents: a comparison of their causal patterns, PhD thesis, University of Strathclyde, Glasgow, U.K.



---

# 32 Ergonomic Interventions in the Implementation of New Technical Systems

*Ken Eason*

## CONTENTS

|                                                                      |     |
|----------------------------------------------------------------------|-----|
| Introduction.....                                                    | 837 |
| Technocentric Implementation Strategies.....                         | 837 |
| Alternative Approaches to the Change Process .....                   | 839 |
| Ergonomic/Human Factors Interventions and the Process of Change..... | 840 |
| Scoping and Assumptions.....                                         | 843 |
| Technical System Implementation Strategies .....                     | 843 |
| ‘Greenfield Site’ Implementation.....                                | 844 |
| Big Bang.....                                                        | 845 |
| Parallel Running.....                                                | 846 |
| Phased Implementation .....                                          | 847 |
| Trials and Dissemination.....                                        | 848 |
| Incremental Implementation .....                                     | 849 |
| Conclusions.....                                                     | 850 |
| References.....                                                      | 851 |

## INTRODUCTION

One of the most frequent sources of change in working life is the implementation of new technology. This can have many implications for the work system. There will be an immediate need for staff to acquire new skills and there may be potentially far-reaching effects on jobs and the way work is organised. There are many opportunities for ergonomic/human factors (E/HF) interventions, but what can be achieved depends on the way the new technology is implemented. This chapter examines the different strategies by which new technology is implemented into the workplace and their implications for E/HF interventions.

## TECHNOCENTRIC IMPLEMENTATION STRATEGIES

The prevailing approach to technical system implementation has been described as ‘technocentric’ or ‘technology push’ (Badham 1993, Eason 1993). In this approach, the objective is seen in entirely technical terms: to design or purchase a complete technical system, to install the equipment, get it ‘up and running’ and train the users in its correct usage. If there is recognition that there may be wider ramifications for the work system, this is often framed as ‘resistance to change’ on the part of the workforce and is to be overcome by such mechanisms as appointing ‘user champions’ to act as persuaders and using industrial relations procedures to deal with disputes.

This approach has been linked to many examples of failed implementations or very slow, painful and costly uptakes of new systems. Balogun and Hope-Hailey (2004) calculate, for example,

that 70% of such change programmes fail and the Standish Group (2014), which has made annual management surveys for the last 30 years, puts the failure rate of major IT implementations in the United States between 30% and 50%. It is notable that the failure rate is highest for ambitious new technology systems where the system is intended to transform or 'modernise' the way a work system operates. There have been a number of high-profile examples in the UK public sector in recent years in which major IT systems have either failed or been very difficult to implement (King and Crewe 2013). Examples that have received considerable publicity include the London Ambulance System (Page et al. 1993), the Passport Office system (NAO 1999) and the computerised fire service control rooms (NAO 2010). In these examples, a technocentric approach was adopted and it was only during implementation that the problems of using the technical system within their host organisations became apparent.

A recent example of a technocentric approach on a large scale has been the National Programme for Information Technology (NPfIT) in the National Health Service (Eason 2005, 2007). This 10-year programme was launched by the Department of Health in 2002 and was intended to get all healthcare agencies across England to share electronic patient data. Hitherto healthcare agencies had implemented their own systems with the result that patient data could not be transferred between them although patients frequently needed treatment by different agencies. The aim of the NPfIT was to replace any existing electronic patient information systems with a standard 'best of breed' application so that everybody could share information. The NPfIT included a number of system developments, but the primary one was the Detailed Care Record System (DCRS) which was intended to enable the complete 'cradle-to-grave' record of a patient to be shared between general practitioner clinics, hospitals and other health agencies. The DCRS was chosen centrally and was to be 'rolled out' across all the health agencies in England, but it immediately ran into major implementation problems. The National Audit Office (NAO) that evaluates the use of public funds in the United Kingdom concluded in an evaluation in 2008 that 'the plan to deliver detailed care records was 4 years behind schedule' (NAO 2008) and in 2011 that 'the plan was unachievable' (NAO 2011). One of the oft-stated problems was that 'one size does not fit all' (Ritter 2008), that is, a single database could not meet the needs of all the different healthcare agencies. Attempts to customise the technical system to meet local needs and deal with all the changes in healthcare work practice that were necessary to make successful use of the new system proved time consuming and difficult. The Royal Free Hospital in London, for example, estimated that it cost £7.2 million to implement their DCRS (Hoeksma 2008). Across the country the mismatch between what the systems provided and what the healthcare agencies needed led to many declining to implement the national system they were offered. As a result, the NPfIT began to deploy a broader range of systems and, towards the end of the programme, local agencies were making their own decisions about the systems they would implement. This approach has now become national policy following the abandonment of the NPfIT. The positive outcome is that local developments can ensure that technical systems meet local needs, but the danger is that the original objective of sharing electronic records across the nation will be lost because the separate systems will be unable to communicate with one another. New technical systems developments often encounter this dilemma: a need to standardize across an organisation and also a need to meet many different local needs. Implementation strategies are needed that provide opportunities to address this issue.

This example demonstrates that the problems of implementation are not just about people resisting change. The problem is that a work system is a socio-technical system (see Chapter 28) in which work is done by tightly interdependent social and technical systems. A change in one part of the technical system has consequences for the other components of the work system. At best, if a technocentric approach is adopted, the other components will need adjusting to operate effectively with the new technical system. At worst, the new technical system will have to be changed before it can be effectively integrated into the work system. When a technocentric approach is adopted, most of these issues only emerge when implementation is attempted and, at that stage, they can be very difficult to address.



## ALTERNATIVE APPROACHES TO THE CHANGE PROCESS

There has been widespread recognition that a technocentric approach to organisational change is likely to lead to implementation problems and there have been many proposals for undertaking change processes in ways that might be more successful:

1. *Continuous rather than discontinuous change*: Many of the formulations of the change process go back to the original work of Lewin (1947) who proposed a model of change in which a system in a steady state is disrupted by a need to change and then regains a new steady state. Lewin also described it as a process of 'un-freezing' (from the old steady state), 'transitioning' (undergoing change) and finally 're-freezing' (into a new and better steady state [Katz and Kahn 1966]). Burnes (2004) suggests that in a modern context this is an outmoded concept because work organisations are now in a constant state of change as they struggle to cope with a rapidly changing environment. He therefore proposes that the way we undertake change needs to be based on continuous rather than discontinuous change. One implication is that, as Bridger (1990) points out, everybody in the work system has a 'double task' to perform; doing the work that is needed today and planning the work system that will be needed tomorrow. Burnes' argument is that if there is a continuing process of dealing with change internally, change will be easier to manage.
2. *Planned change versus emergent change*: A common finding from evaluation studies that examine what happens when new technology is introduced into a work system is that instead of the planned adoption programme there are unexpected emergent changes (Latour 1986, Oudshoorn and Pinch 2003, Abdelnour Nocera 2007). The workforce responds to the new technology by adopting some features, ignoring others and developing workarounds when what the technology requires does not match what the users need to do to complete their tasks. The collective result is that the technology is assimilated into the work system to the extent that the workforce perceive it as valuable, usable and effective. Orlikowski (2000) refers to this as the workforce developing a 'practice lens', that is, seeing the technology in terms of the way it relates to their work practices. Bijker (1995) and Pinch and Bijker (1987) coined the term SCOT, the Social Construction Of Technology, to represent all the ways in which the social system re-construes the technology according to its own needs and culture. If this is what actually happens, the argument is that the implementation process might better consist of providing a technological capability and supporting the workforce as they find the most appropriate way of adopting what is useful. It should be noted that this proposal means that change emerges from within the work system as opposed to the situation described in the NPfIT example earlier where change is the product of external forces. Leucke (2003) defines this as the difference between planned change and evolutionary development, and Bamford and Forrester (2003) as about planned developments as opposed to emergent developments. Change processes that emerge from within are much more likely to be successful because they are 'owned' by the staff of the work system and are based on a detailed understanding of the work that needs to be undertaken.
3. *Organisational learning, adaptation and evolution*: One of the most significant contemporary developments in organisational theory is the emphasis being placed on the adaptive capabilities or resilience of the organisation in the face of continuous environmental change. Senge (1990), for example, explores the many ways organisations can support the continuous learning of all members of staff. Discussing how to develop adaptive enterprises, Haeckel (1999) draws attention to the need for a 'sense and respond' culture at all levels in the organisation: that is, to maintain a way of reviewing what is happening outside the organisation and to have rapid ways of changing the internal work system in order to respond to opportunities and threats.

These formulations of organisational change suggest that success comes from the workforce sensing the need for change and finding ways of using new technology to adapt the work system in order to respond. And this needs to be done on a continuous, probably incremental, basis. In this way the relevance of the change to the whole system is never lost and there is strong commitment to the change from the workforce. However, taken to its logical conclusions, this approach would have every work system in a large organisation developing in its own way. This could lead to a very uncoordinated organisation. If an organisation wishes to implement a major policy change, for example changing its products to fit an emerging market, it has to implement a coordinated set of new systems in many parts of the organisation. There has therefore to be some degree of ‘top down’ implementation which may be experienced by front-line staff as imposition of change from outside. How is it possible in a change process to have both external imposition and internal control over the change process? This issue has been a preoccupation of socio-technical systems theorists for many years, and Cherns (1976) expressed the target in the form of the principle of minimum critical specification, a requirement, for example, that the external specification for new technology is limited to the level necessary for coordination across the organisation in order that there is scope for local staff to customise it for use in their work system. Eason (2007) provides a case study in which healthcare staff were able to shape an electronic patient information system to meet the needs of mental health patients because there was some flexibility in the design of the technical system.

## ERGONOMIC/HUMAN FACTORS INTERVENTIONS AND THE PROCESS OF CHANGE

There are many purposes an ergonomist might have when making a contribution to the implementation of new technology into a work system. If the implementation is undertaken within a technocentric approach, the interventions that are possible may be framed by the technology rather than the broader work system, for example, may be about interface and workstation design for the new technology and training staff in the skills needed to operate the technology. However, if there is recognition in the approach to change that there will be consequences for the work system and that socio-technical work has to be done to assimilate the new technology into the system, the opportunities for interventions become greater. Further, if there is recognition that the change process must involve the workforce in a long-term learning and adaptation process, there are also separate issues about the immediate goals of an intervention and the more enduring goals that might be sought. In Table 32.1, four possible interventions are identified which lead to specific aims with respect to the design of the immediate work system and to longer-term aims to support the enduring needs of the work system.

1. *Socio-technical systems integration*: A new technical system has many implications for social sub-systems that have to be addressed if an integrated system is to result (Eason 1996). In the short term, this often means considering issues of human–machine interaction

**TABLE 32.1**  
**Immediate and Enduring Aims of Ergonomic/Human Factors Interventions**

| Intervention                        | Immediate Aim                                                 | Enduring Aim                                                                         |
|-------------------------------------|---------------------------------------------------------------|--------------------------------------------------------------------------------------|
| Socio-technical systems integration | To create effective forms of human–machine interaction        | To sustain an effective socio-technical system                                       |
| Safety and health                   | To design work settings that are safe to work in              | To create a safe and healthy work culture                                            |
| Learning and adaptation             | To develop competence for current work goals                  | To sustain the work system as an adaptive, learning organisation                     |
| Motivation and participation        | To engage the workforce in developing the current work system | To establish jobs and forms of empowerment that promote enduring forms of motivation |

to ensure the new technical system has the functionality and usability necessary for the workforce to achieve their tasks (see Chapter 13 for examples in human-computer interaction). In the long term, the technical system and social system may need ongoing adjustments in, for example, job structures, recruitment, training, reward systems or technical support to sustain integration.

2. *Safety and health*: A new technical system can raise new questions about the safety and health of the workforce both directly, for example in the operation of machinery, and indirectly, for example in the changing nature of work patterns or the creation of new physical work environments. In the short term, an E/HF intervention may consist of a risk assessment of the new technical system and the design of suitable work environments and work practices to protect safety and health (see Chapter 30). In the longer term, there will be continuing changes to the work system and, if we assume an ergonomist will not be in continuous attendance, there will be a need to establish a safety and health culture which can review and adapt safety practice.
3. *Learning and adaptation*: A new technical system may require the workforce to learn new skills to operate it and there may also be new work practices to master. The development and use of relevant training techniques when the system is implemented will be a necessary intervention. However, in a world of continuous change, individual learning needs to be continuous and the work system has to be in a process of continuous adaptation to newly emerging needs and opportunities. Meeting not just the immediate need for new skills but the enduring need to keep refreshing the competence of the workforce is a major challenge for an E/HF intervention.
4. *Motivation and participation*: It is widely acknowledged that technical change is much more successful when employees are motivated to accept it. Two related routes are advocated. First, the workforce can participate in the change process, with real power to influence events and they can then develop a sense of ownership of the change (Damodaran 1996; see Chapter 34 of this book). Second, by designing work with new technology so that it meets requirements for satisfying, worthwhile jobs, people will have something tangible to gain from the change process. However, people continue to develop in their work and careers and the process of change is continuous. The enduring need is not therefore a one-stage participation to produce better jobs but the establishment of ongoing participative mechanisms to support the long-term empowerment of the workforce in the design of the work system of which they are a part.

An example of a development where attempts are made to deal with all of these immediate and enduring needs is the introduction of lean production methods into manufacturing plants approach that originated in Toyota. Lean production involves major technical and process changes, but attention is also given to the immediate implications for the jobs of the work force. However, it is also an approach that recognises the need for continuous change and quality circles are established at shop floor level to ensure the workforce can participate in the quest for continuous improvement in the years following the implementation of lean production (Koukoulaki 2014).

One feature that is common across all four kinds of intervention is the difference between meeting immediate needs and creating a work system that is adaptive in the longer term. In the short term, E/HF interventions may ensure the new socio-technical system created when the new technical system is implemented is effective, safe and healthy for the workforce to engage in current operations. In the longer term, the aim is to improve the capability of the work system by enhancing its adaptive mechanisms so that it can cope with future demands for change. In a survey of social science and ergonomic interventions made in organisations in the United Kingdom and Germany, Klein and Eason (1991) found this distinction to be of concern to many practitioners of the human and social sciences. Many of them were given the opportunity to make contributions to the current system within limited timescales and terms of reference. They were worried that their work would

be transient, swept away when the next wave of change came along. Many of the practitioners were working to establish practices in the organisation that would sustain good working practice in the longer term. Klein and Eason call this 'institutionalising' the practice so that it becomes part of normal procedure within the organisation. A good example is the 'devolving to ergonomics' approach (Wilson 1994). An ergonomist might, for example, be invited to design the equipment, interface or workstation associated with a new technical system. If the ergonomist uses the opportunity to train some of the workforce in the rudiments of activities such as auditing workstations, these members of staff may be able to monitor future circumstances and trigger actions when necessary to sustain good ergonomics/human factors (E/HF) practice.

Underlying these methods of making E/HF interventions, especially those aimed at the long-term capability of the work system, are mechanisms for supporting the engagement of the workforce in the change process. Four such mechanisms have been found to be particularly significant in successful change processes:

1. *Participation*: Fundamental to the achievement of successful change is the participation of the workforce in the change process. The methods that can be used are fully explored in Chapter 34. In relation to the implementation of new technology, the purpose of participation is particularly important. It has to have a focus on relevant design for the people concerned: it has to be about the design of the technology to meet the needs of each section of the workforce and it has to be concerned with the (re)-design of their work practices and their jobs.
2. *Change agents and facilitators*: A common practice in change programmes is to appoint one or more of the local staff to be change agents or facilitators: to take responsibility for enabling the change to take place in that locality (Akrich 1995, Hall and Hord 2006). With respect to the implementation of new technology change, agents can act as an important communication channel between technical system designers and the workforce. In the health service, for example, 'clinical facilitators' are established who act as go-betweens in technical change processes, interpreting the needs to healthcare staff to technical design staff and helping healthcare staff see the implications for their working practice of whatever technical solutions are proposed (Eason et al. 2012). Providing change agents with methods and knowledge can be a very powerful way for an ergonomist to create long-lasting change in an organisation.
3. *User-centred design*: A number of authors have identified the way users need to be involved in the different stages of a design process (Galer et al. 1992, Eason 1995, Damodaran 1996) and to the need to involve all the different stakeholders in the process. Different techniques, for example, are needed to help users articulate their requirements at the beginning of the process, to enable them to assess the implications of technical system proposals and to support them as a new system is being implemented.
4. *Review mechanisms, audits and evaluations*: Authors who focus on creating learning mechanisms in organisations point to the need for regular review processes where members of the workforce can reflect on how things are working and make proposals for improvements. In respect of new technology implementations, it is often the case that uptake of the technology lags behind the installation of equipment because it takes time for people to assimilate the new capability into their normal working practices. The speed of technical change and the rapid obsolescence of the technology already in use means that people at work are frequently faced with the challenge of adjusting to new forms of technology. Undertaking regular audits of usage and problems after installation can make opportunities and issues explicit and enable people to share in the development of emergent new practices. In many cases, there may be formative evaluations during implementation (that can have a direct influence on the technical system being implemented) and summative evaluations after the new system is in place (that can facilitate the adoption of new working practices).

## SCOPING AND ASSUMPTIONS

Every opportunity to make an ergonomic intervention is different with its own constraints and challenges. The initial stage of an intervention needs therefore to include a form of scoping to see what can and cannot be done. Operating within an open socio-technical systems perspective, Eason et al. (1996) and others have developed the organisational requirements definition for information technology systems (ORDIT) methodology to help the staff of work organisations plan changes in their work systems as a result of the implementation of new information technology systems. The first stage in the ORDIT methodology is scoping, and it involves four activities.

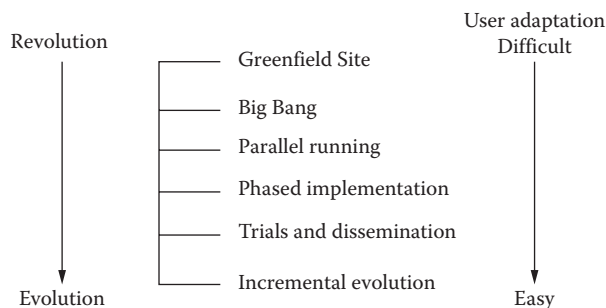
1. *Analysis of the current system*: This is an outline of the current socio-technical system which is likely to be affected by planned changes. This analysis helps to establish the boundaries to the system and to get a perspective on history, work culture and current work practices.
2. *The planned change*: This is an assessment of the drivers for change (what objectives are being sought), the form of the change (e.g. it might be a change in the technical system) and the state of development of the plan.
3. *Key stakeholders and principal requirements*: An analysis of the stakeholders within the work system who are likely to be affected by any changes, for example the management who commissioned the change, the workforce who will use any future technical system and the technical staff who are implementing it. For each group of stakeholders, it is useful to identify key requirements in relation to the change. The managers may, for example, expect improvements in system performance. The workforce may have more defensive requirements, for example to keep their jobs or to sustain work practices.
4. *The contract for the intervention – explicit and tacit*: The practitioners will have been commissioned to play some role in the technical system implementation and scoping is a phase when this can be explored, made more explicit and to some extent negotiated. The request may, for example, be to contribute on safety and health, but it may be desirable to extend this to consider aspects of systems integration, for example how safety relates to incentives or to training, and to explore what opportunities exist to pursue enduring aims as well as contributing to the immediate changes to be made.

For the purposes of this chapter, it is assumed that scoping has shown that a major technical system is to be implemented and that there is recognition among the stakeholders that this could have significant ramifications for work practices and jobs. The process of implementation will not therefore be entirely technocentric and will have scope to deal with changes in the socio-technical system. The scoping process will also have revealed the implementation strategy that is to be used, and it is likely to be one of the six discussed later (or some combination of them).

## TECHNICAL SYSTEM IMPLEMENTATION STRATEGIES

The six major strategies used for technical system implementation listed in Figure 32.1 are derived from Eason's (1988) examination of the strategies used in the implementation of information technology systems. The strategies are arranged on a dimension from the most revolutionary to the most evolutionary, that is, in Burnes' (2004) terms from discontinuous change to continuous change. At one extreme, we have the major start-up of a complete new system overnight and, at the other, the steady introduction of technical facilities over an extended period of time. There are two features of this dimension that are particularly important. First, the more revolutionary approaches involve the greatest risk to system performance because so many components of the system are being changed at the same time. Second, as discussed earlier, it is likely that the workforce will have more opportunities to understand the implications of the new technical system and to influence the way the system is adopted if the change process is evolutionary. It will be more difficult for them to adapt and





**FIGURE 32.1** Implementation strategies.

influence system changes if the strategy involves sudden change. However, there are many situations where radical change is necessary and ways of addressing the wider system ramifications have to be found if the change is to be successful.

The aim in this section is to identify the main principles of each strategy and the circumstances in which it might be used. We will then consider the implications for E/HF interventions.

### **‘GREENFIELD SITE’ IMPLEMENTATION**

This is when a new technical system is part of a complete new start-up, that is, a new factory is built, a new shop is opened. The work process and the technical system may be based on existing systems elsewhere, but designers have a chance to introduce new processes and technologies. Once the new system is in place, there is often a lengthy commissioning period when the whole process is tested before the factory produces products or the shop opens for business (Emery 1980). Although the ‘Greenfield Site’ approach provides opportunities for testing before ‘go live’, there are often unforeseen problems when the new system has to operate at full scale in a real environment as was found when the new Terminal Five building opened at Heathrow in 2008 and immediately ran into major baggage-handling problems. The subsequent inquiry demonstrated the risks in the start-up of a major new system if testing is not done at full load and with all of the intended socio-technical system in place (House of Commons Transport Select Committee 2008).

From one perspective, a the Greenfield Site is an excellent opportunity for an E/HF intervention. Many features of the work system are being newly created, and it may be possible to pursue both immediate and enduring aims listed in Table 32.1. There could be opportunities to work with the designers of the new process on many issues, on elements such as workstation layout or the physical environment, to ensure good human–machine interaction and safe and comfortable working conditions. It may also be possible to apply many of the principles of job design (Alberg 1981, Parker and Wall 1998) to create work with meaning, variety and feedback that will be satisfying and motivating for the workforce. Klein and Eason (1991) give an example of the development of a new chocolate factory in which the social system was based on autonomous work groups and planned on the basis of job design principles.

From another perspective, however, the Greenfield Site is a problem for E/HF interventions because the workforce may not be appointed in time to play a part in the planning process. Typically, the workforce will be newly recruited in time for training before the new system goes live. Decisions may need to be taken about the jobs, work processes and the workstations that the workforce will encounter. Designers may, of course, have the best of intentions with respect to the health and welfare of the workforce and may have access to good ergonomic advice. However, there may be little opportunity to engage the workforce in the planning process and begin the process of building a learning organisation in which staff ‘own’ the work system and seek continuous improvements in its performance. For these purposes, it is better to recruit staff early and/or leave many of



the decisions about working practices and organisational procedures until later, that is, adopt the principle of 'minimum critical specification', so that the workforce can play their part in the planning process and develop a sense of ownership of the new system. The literature on the planning of change emphasises the importance of providing the workforce with these opportunities to influence the working practices they will have to employ. As will be seen in the discussion of the strategies later, this is common when the workforce already exists. When it does not exist, the desire to make all the decisions for future staff has to be resisted if the long-term aims of an effective intervention are to be realised.

## **BIG BANG**

One of the most difficult kinds of implementations to undertake effectively is when an existing work system is discontinued in its entirety and replaced immediately with a new system. This usually happens when an organisation needs to sustain the continuity of its business but, because of system interdependencies, has to change a large 'critical mass' of activities to bring in innovations. One of the most well-documented examples was the overnight switch of the London Stock Market to electronic trading in 1986. Stock market trading involves complex interactions by many trading organisations, and it was necessary for everyone to make the switch at the same time. The London Stock Market 'big bang' involved the implementation of a completely new technical system that enabled traders to do deals from their offices rather than on the trading floor. It also involved a major social change in which the stock market trading was opened up to many different agencies and not just those with access to the trading floor. Obviously making major technical and social changes on this scale can be a high-risk strategy because all the parts of the new system have to function together immediately if the normal level of business is to be sustained. This is difficult to achieve and often there is an 'initial dip' in performance before the gains expected of the new system are achieved. In the case of the London Stock Exchange, everybody was so aware of the dangers to trading if the new system did not work effectively that planning and testing were very thorough. In the event, while there were minor hiccups, the transition went well with only a small and temporary 'initial dip'.

But quite often the dip can be deep and long. When the UK Passport Office switched to a new passport issuing system in 1999 (NAO 1999), performance dropped to 27% of the previous issue rate and did not recover for 18 months. Again a new IT system underpinned the change, but a major factor in the problems encountered was the splitting of the work between three organisations. Routine work was 'outsourced' to two private companies while sensitive and complex tasks were kept within the Passport Office. An underestimate of the volume of complex work compared with routine work led to the new division of labour between the organisations not coping with the workload that was experienced.

Some of the problems in 'big bang' implementations may be technical because it may be the first time the complete system has been tested under a full load. The London Ambulance Service introduced a new computer-based ambulance despatch system in 1992 (Page et al. 1993, Wastell and Cooper 1996). The aim was to use artificial intelligence software to decide which ambulance to despatch to which incident and which hospital it should take patients to. In the event, the technical system could not cope with the nature and number of calls made to it and there were many instances of patients waiting many hours for an ambulance. The system was withdrawn and was never reinstalled.

If a major change of this kind is undertaken very quickly without full preparations, there is a high probability that there will be a failure in some parts of the new socio-technical system. A particular problem is the strain it may place upon the workforce who may find themselves with new job demands and unexpected problems to solve. Time and opportunities need to be built into the period before 'go live' for the workforce to participate in systems development and to prepare themselves for new work practices. If there is acceptance of the need for detailed participative planning before the change, there should be many opportunities to ensure good job design and examine health and

safety issues. As the new system takes shape, it may be desirable to organise large-scale, 'offline' 'dry-runs' or rehearsals of the new ways of working to test that everything is in place and to give staff a chance to experience what is to come. These events often enable staff to identify unintended consequences of change, for example where there are ramifications for people who are not directly affected by the change but who find their working practices disrupted by the changes in the work that others do.

When it is well managed, there are advantages to a big bang implementation; it becomes the big event and, if everybody is committed to its success and shares in the achievement, it can be a major shared experience for the organisation. The collective energy and excitement of such an event can carry the new system through many teething troubles and create lasting bonds between members of the workforce. It may well become an important part of the history of the organisation, spoken of with affection by those who lived through it.

The one danger of a successful instant changeover of this kind is that it will be seen as heralding a period of stability (or in Lewin's terms a period of 'refreezing') when there is no further need for developmental activities. And yet there will be a continuing need for systems adaptation and for further learning and staff development. One way of facilitating these enduring goals is to keep in place some of the participative structures created for the planning of the change. They can, for example, take responsibility for summative evaluations of various attributes of the new system from productivity to safety and health.

## PARALLEL RUNNING

A popular way of minimising the risks to business productivity when introducing a new technical system, especially if it is a computer-based system, is to run the new system alongside the old system. When it is 'mission critical' that no loss of quantity or quality in output should occur during the system change, this can be a secure way of proceeding. In socio-technical systems terms, what happens in this form of implementation is that the social system operates two technical systems at the same time. The old one is used to deliver products and services while the work is repeated with the new system in order to detect and remove any technical problems and develop the competence and confidence of staff in using the new system.

As a process of change, this approach can have positive effects on management, the workforce and designers. It gives everybody the time to test the new system in a field setting without having to rely on it. It is not, however, without its drawbacks. Operating both systems, for example, involves extra work and it is rarely the case that extra human resources are available. If the new system keeps failing, there may be a reaction against the extra work and a desire to retain tried and trusted methods. On the other hand, if the new system works well and saves a lot of time and energy, it can be frustrating to continue using the old system. In the freight-forwarding case study described in Klein and Eason (1991), a new system provided a way of capturing data once and then using that data for a wide range of clerical functions. The old system required the users to repeat the data capture for every clerical function. After a short period of parallel running, the staff pleaded to be allowed to use the new system for real tasks.

Within the parallel running approach, there is an opportunity to work with the workforce and the designers on the interface issues and the workstations for the new technical system. What is more problematic is to work on any job or organisational changes that might be appropriate when the new system is fully in place. The continued operation of the old system requires the existing organisational structures and procedures to remain in place. This can reinforce organisational inertia to keep things as they are, and this may mean that the opportunity to change the entire socio-technical system for the benefit of the staff and the business is missed. However, parallel running does provide an interesting opportunity to work on sociotechnical opportunities. A number of authors have advocated a participative approach to considering work organisation and job design questions when technical change is planned, for example Mumford (1983, 2000) and Eason and Sell (1981).

One difficulty with these approaches is that often the workforce does not have sufficient understanding of the new technical system to be able to examine its consequences for alternative forms of work organisation. In parallel running, the workforce develop a detailed knowledge of the technical system and are well able to participate in job design discussions. Planning organisational change to take place at the point when the technical system has been fully commissioned can be an effective way of ending a period of parallel running.

### **PHASED IMPLEMENTATION**

The risks associated with making massive changes in a work system at one point in time can be eased by phasing in the changes over a period of time. There are two ways in which large-scale changes can be sub-divided to facilitate phased introduction. First, the functionality of the technical system can be introduced in phases so that, for example, basic facilities can be introduced first and more advanced ones at a later date. Second, it may be possible to introduce the system in different parts of the organisation at different times. It is this latter approach, often called 'roll out', which will be examined in this section. Phased implementation is an attractive option for organisations that have many semi-autonomous sections, for example branches of a bank, building society or government agency, or a chain of shops. If a similar technical system is to be implemented across the organisation it can be done as a series of 'little bangs' taking each branch or section at a time. The disturbance of the change is localised by this strategy and has little impact on the overall productivity of the organisation. Another advantage is that an implementation team can be assembled with all the necessary expertise and they can give each branch their full attention for a short period to introduce this new system with maximum efficiency.

This approach has many advantages if the aim is to make ergonomic interventions. If ergonomic competence is built into the multi-disciplinary implementation team, there will be opportunities to influence the new forms of human-machine interaction, workstations, the physical environment and training. Since these activities are repeated in each branch, there will be an opportunity to influence these activities across the whole organisation. As with the other strategies there are dangers. A 'roll out' of a technical system of this kind is most often a centrally planned process in which the new system is largely imposed on each branch. There is often an imperative to get the system implemented as quickly as possible and therefore little opportunity for the workforce at each branch to do more than learn the new work practices. And yet for each workforce this 'little bang' is just as much upheaval, uncertainty and rejection of current practices as it is for the workforce in the 'big bang' implementation. This can lead to greater resistance as the roll out proceeds. As Tapscott (1982) observes in his 'second law of office system automation', the 'difficulty of implementation increases with every new implementation'. This is often counterintuitive to the implementation team who, having solved the technical problems and worked out efficient procedures for the changeover, may expect each new implementation to be faster and smoother. For the workforce, however, it can seem like the swift imposition of procedures established elsewhere with little attention to their local conditions or practices.

One technique which can avoid these outcomes is to ensure Cherns' (1976) 'principle of minimum critical specification' is applied during central planning, that is, only the features of the system that have to be standardised to permit inter-branch cooperation and control are standardised. This can leave substantial discretion for local variations to meet local conditions. Individual branches or units typically differ from one another in significant ways, for example in customer or caseload, in geographical area, in premises and in staff profiles. There may be quite substantial and legitimate variations in work practice to accommodate these differences. If the new technical system allows some flexibility, it is possible for local staff to work out the specific ways in which they will adopt the new system, for example personalising some of the interfaces, establishing local job design or matching workstation layouts to the premises. Where these opportunities exist, a local working party can be convened ahead of implementation to plan how the implementation can

best proceed for their branch. An ergonomist can play a key role in helping the workforce exploit system flexibility so that the system meets their needs and they develop a sense of ownership of it. Eason (1995), for example, describes the implementation of an office automation system into a number of departments in a legal practice. In each department, the clerical staff planned specific forms of training, chose their own furniture and established their own workstation layout. These were important areas of participation and influence for the staff and contributed to the rapid adoption of the new system.

## TRIALS AND DISSEMINATION

A common way to initiate a major organisational change is to make a trial or pilot implementation and, if this is successful, to disseminate the resultant system to the rest of the organisation in a 'roll out' programme similar to the phased implementation approach described earlier. This differs from a prototype test or an offline simulation by being a live running of an intended new system. In some respect, the trial is like the first implementation of a phased implementation programme except that the system is under test and there could be major changes before full dissemination. If the concept of a trial has real significance, it might mean, for example, that, following a review, a decision could be taken to abort dissemination or undertake a different kind of development. Trials or pilot implementations are often undertaken in major organisations that have, for example, many branches or who have critical performance standards to sustain and have to reduce as many risks as possible before major commitment.

The fact that this is explicitly a learning strategy makes it ideal as a vehicle for E/HF interventions. As in the other examples, the fact that it is a live implementation means it is an opportunity to test whether forms of human-machine interaction are fit for purpose and usable by the workforce, and that the physical workstations conform to health and safety requirements. The staff themselves can, if possible, play a part in the detailed matching of the technical system to their particular circumstances. However, the formal fact of a trial raises other possibilities. A technical system embedded in a work process and interacting with the existing culture of the social system will have many implications; for jobs, for work practice and for rewards, power and influence. Many of these issues may be unsuspected by those creating the new technical system, but the trial can make them evident. An evaluation of the wider organisational consequences of the technical system can often turn what was expected to be a proving trial for a technical system into a socio-technical systems reappraisal. In the freight-forwarding system cited earlier (Klein and Eason 1991), a trial implementation of the new technical system was made involving three neighbouring branches that were to use the system as a common database to consolidate loads before export. The trial demonstrated that while the database was extremely useful to the clerks, its use for load consolidation raised many questions for management; 'whose load was it' or 'who got what share of the revenue'? After the trial, the organisation went ahead with a branch-based database system but discontinued the idea of load consolidation because it raised too many organisational issues.

There are three common issues which affect the usefulness of a trial implementation from the perspective of ergonomic interventions:

- *Ensuring the learning:* At its best, the trial is a strong evidence base for action learning (Revans 1982), that is, it constitutes experimental actions from which members of the organisation can learn and improve performance. However, it is often the case that lip service only is paid to the trial aspect of the implementation. If it is treated as a technical system proving trial and if are few problems, there is a tendency to proceed quickly to full dissemination without reviewing the implications for the workforce or organisational practice. Often the dates for the implementation 'roll out' have already been pencilled in. If the trial includes two features, this outcome can be avoided. First, an

agreed programme of evaluation is needed in which data will be gathered systematically about performance, usability and organisational implications. Second, the body assessing the trial, preferably including management, design team and representatives of the workforce, needs terms of reference to receive and comment on the evaluation before dissemination can continue.

- *Recognising the 'Hawthorne effect'*: It is often the case that a trial implementation goes very well, lulling the organisation into a false sense of security and there is a rude awakening when dissemination in the rest of the organisation proves much more difficult. Two factors can cause this to occur. First, the organisation may have chosen a part of the organisation for the trial that fits the technical system well and has an amenable workforce. A more useful test would have been to do the opposite, but it would be a very brave project leader who would take this risk. The second issue is that a trial is tailor-made for the 'Hawthorne effect' (Rothlisberger and Dickson 1939 and see Chapter 4). This is the well-known situation where a change proves very effective in terms of productivity not because of the change itself but because the workers find themselves 'in the spotlight' and respond to the extra attention they are receiving. Being chosen for a trial can be exciting or nerve wracking; either way it is not an ordinary occasion. Senior managers and specialists may be very much more in evidence than normal. The workforce may also find themselves in the novel situation of being asked for their opinions and changes may be made on the basis of their views. All of these factors may encourage strong motivation and extra effort. None of these factors may be present in later implementations when it has all been done before and no one is interested in the views of the staff. This is not an argument for avoiding the 'Hawthorne effect'. Rather it is a statement that it will happen, that its effects need to be recognised and, if possible, reproduced in later dissemination.
- *Preparing a dissemination strategy*: The trial needs to be used to plan dissemination procedures. There are some obvious ways of doing this, for example using it to establish who needs to be trained in what ways. However, the trial can also be used to start the process of involving the rest of the workforce. In one major government department, for example, the normal procedure is for user representatives from around the country to be invited to the trial site to talk to their colleagues about the experience of the new system. They are then asked to start planning for their own local implementation. If there is some freedom to adopt the system for local needs and practices, there will be a real opportunity for local design and ownership which will both improve the integration of the socio-technical system and avoid the dangers of resistance to change.

## INCREMENTAL IMPLEMENTATION

The final implementation strategy is one that is becoming increasingly common as generic forms of technology become widely available as consumer products. Whereas the kinds of technical systems involved in 'Greenfield Sites' or 'Big Bang' implementations are often custom built, or at least put together in unique ways, it is now quite easy for organisations to purchase standard pieces of technology for rapid installation. This is very common in the information and communication technology industry where 'shrink wrapped' products, from new software packages, PCs to telephone equipment, can be acquired and implemented without a lengthy development process. The effect is that, over a period of time, an organisation can acquire a sophisticated set of technical facilities through a series of incremental steps without having engaged in an organisation-wide process of technical change.

For the workforce, there are many advantages to this form of technical capability acquisition. First, it is likely to be a distinctly user-centric process. It may well be that it is the individual or a work team that is choosing the technology to meet a particular need. Even if this is not the case, it is quite likely that the workforce has a lot of discretion as to whether they do or do not make use of



the facilities on offer. In incremental implementation, the new facility is often a specific upgrade, an addition to technology already available and may not be essential to the conduct of the job. Another gain for the workforce is that there is no major learning to be done and no major changes in the work process.

Ethnographic studies have demonstrated that each work system develops a unique cultural microclimate and workforces evolve particular work practices that work in their context. Hutchins (1995), for example, studied the particular work practices of a team responsible for docking commercial shipping, and Brown and Duguid (2000) report the way photocopier maintenance staff shared knowledge about the problems they encountered with each new product. If each workforce can be responsible for selecting their own new technology, it should enable them to sustain useful work practices and use the technology to further develop their capability.

If this is truly a user-controlled, consumer-oriented approach to technical change, maybe there is no need for an ergonomic intervention. As long as the human factors issues of the generic products are attended to by the supplier perhaps, there is no need for attention at the point of implementation? Unfortunately, the evidence suggests otherwise. The piecemeal and localised development of technical capability can cause many problems. The incremental addition of equipment to a workstation can, for example, turn what might have been a well-planned work base into something disorganised and possibly dangerous. The workforce or its advisors may not choose equipment in a planful way and may find it not usable or fit for function. They may also underestimate the amount of learning required. Bajer (1998), in a study of software upgrade implementations, for example, found that most staff assumed the new system would work in the same way as the old one and only after many errors did they discover they had to develop new skills. The cumulative effect of these factors and the discretionary nature of the users can mean that very little of the new functionality gets used.

Another set of problems arising from this piecemeal approach is how to sustain a cohesive socio-technical system. Technical support staff often find, for example, that it is difficult to support the wide variety of technology being acquired, and it is difficult to sustain interoperability between different kinds of equipment. Similarly, the organisation may become less cohesive as each section develops its own practices. These factors suggest that there is a real need for ergonomic interventions in the incremental approach to technical change. However, because it is fragmented over time and across the organisation, there is often no clear structure within which an ergonomist can operate. One strategy that we have used in a number of projects with this character (e.g. Eason et al. 2000) is an action research approach. In this approach, localised and incremental developments are encouraged but there is an overarching steering body on which user communities are represented. A programme of regular evaluations is conducted to identify the positive and negative outcomes of the changes being made, and the steering committee has responsibility for reviewing the impact on the integrity of the entire work system. They then undertake the work necessary to ensure the necessary degree of cohesion, where possible, without infringing the autonomy of local units. This strategy provides the opportunity for ergonomists to work with the steering committee and to contribute to the ongoing evaluations. They are then able to assist with implementation issues and, because the staff are already engaged in both design and review activities, use the steering committee as a forum for establishing a continuous learning process supporting longer-term work system objectives.

## CONCLUSIONS

There are many human and organisational issues associated with technical change, and there are a wide variety of contributions that E/HF can make. How that contribution can best be made depends upon the organisational context, how ready the organisation is to recognise that a technocentric approach will not be sufficient and the strategy by which the technical system is being implemented. To be successful, the ergonomist has to 'start where the organisation is' and work



with the leverage and opportunities that it provides. Three overriding themes emerge from this analysis that influence how this is to be done.

- *Engaging with the workforce:* It is fundamental to E/HF that practitioners work with the people who will use the equipment and occupy the workstations. The different implementation strategies provide different opportunities not only to work with the workforce but to trial systems so that changes can be made before implementation. In the revolutionary strategies, this engagement may be problematic either because the workforce does not exist as in the Greenfield Site or because there may be a very limited time period for the engagement, for example, in the Big Bang or phased implementation strategies. It becomes important in these circumstances either to create the opportunity for this engagement in the early stages of the development or to defer detailed design until local decisions can be taken.
- *The participation of the workforce:* One of the aims in Table 32.1 was to create motivation in the workforce to avoid resistance to change and get a commitment to the new system. The aim is to avoid the 'them imposing on us' view and develop a 'this is what we want' attitude. This can only be achieved by giving the workforce the opportunity to influence what is being implemented. This is very much the case with the incremental evolution strategy which tends to be user led. Trials and dissemination and parallel running also give some opportunities for the workforce to be influential. There is a danger that in the strategies that involve sudden change, for example Greenfield Sites and the Big Bang, most of the workforce could be marginalised and resistance to change could result. There are ways of getting leverage for workforce participation in these strategies, but they need considerable commitment and planning in the period leading up to the change.
- *Working for long-term institutionalisation:* It is often the case that ergonomic interventions are made in the short term and that the expectation is that they will contribute directly to the design of the system to be implemented. However, if the long-term aim of creating an adaptive, learning system is to be pursued, part of the work must be to create the learning and reviewing mechanisms that can foster adaptation and transfer useful knowledge to staff. The trial and dissemination approach has a review mechanism built into it, but most strategies do not. Again the requirement will be to look for opportunities to create and perpetuate mechanisms by which staff can engage with developments and use them as learning opportunities wherever possible.

The context of a technical change can be very influential in determining what kind of E/HF intervention is necessary or possible. However, this is not to say that the ergonomist has to be a passive recipient of the constraints inherent in any particular situation. If both the short- and long-term aims are to be pursued, it is likely that the ergonomist will have to be proactive and imaginative to create the conditions necessary for a successful intervention.

## REFERENCES

- Abdelnour Nocera, J.L. (2007). *The Social Construction of Usefulness: An Intercultural Study of Producers and Users of a Business Information System*. Saarbrücken, Germany: VDM Verlag Dr. Mueller.
- Akrich, M. (1995). User representations: Practices, methods and sociology. In Rip, A., Misa, T.J. and Schot, J. (eds.), *Managing Technology in Society: The Approach of Constructive Technology Assessment*. London, U.K.: Pinter Publishers.
- Alberg, U. (1981). Techniques in redesigning routine work. In Corlett, E.N. and Richardson, J. (eds.), *Stress, Work Design and Productivity*. Chichester, U.K.: John Wiley, pp. 157–164.
- Badham, R. (1993). Introduction: New technology and the implementation process. *International Journal of Human Factors in Manufacturing* 3(1), 3–13.
- Bajer, J. (1998). An investigation of the human costs of software upgrades in organisations. Doctoral thesis, Loughborough University, Loughborough, U.K.

- Balogun, J. and Hope-Hailey, V. (2004). *Exploring Strategic Change*, 2nd edn. London, U.K.: Prentice Hall.
- Bamford, D.R. and Forrester, P.L. (2003). Managing planned and emergent change within an operations management environment. *International Journal of Operations and Production Management* **23**(5), 546–564.
- Bijker, W.E. (1995). *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press.
- Bridger, H. (1990). Courses and working conferences as transitional learning institutions. In Trist, E. and Murray, H. (eds.), *The Social Engagement of Social Sciences*, vol. 1. Philadelphia, PA: University of Pennsylvania Press, pp. 221–245.
- Brown, J. and Duguid, P. (2000). *The Social Life of Information*. Boston, MA: Harvard Business School Press.
- Burnes, B. (2004). *Managing Change: A Strategic Approach to Organisational Change*, 4th edn. Harlow, U.K.: Prentice Hall.
- Cherns, A.B. (1976). The principles of socio-technical design. *Human Relations* **28**, 783–792.
- Damodaran, L. (1996). User involvement in the systems design process – a practical guide for users. *Behaviour and Information Technology* **15**, 363–377.
- Eason, K.D. (1988). *Information Technology and Organisational Change*. London, U.K.: Taylor & Francis Group.
- Eason, K.D. (1993). Gaining user and organisational acceptance for advanced information systems. In Masser, I. and Onsrud, H.J. (eds.), *Diffusion and Use of Geographic Information Technologies*. Dordrecht, the Netherlands: Kluwer, pp. 27–44.
- Eason, K.D. (1995). User-centred design: For users or by users? *Ergonomics* **38**, 1667–1673.
- Eason, K.D. (1996). Understanding the organisational ramifications of implementing information technology systems. In Helander, M.G., Landauer, T.K. and Prabhu, P.V. (eds.), *Handbook of Human-Computer Interaction*. Amsterdam, the Netherlands: Elsevier, pp. 1475–1495.
- Eason, K.D. (2005). Exploiting the potential of the NPfIT: A local design approach. *British Journal of Healthcare Computing & Information Management* **22**(7), 14–16.
- Eason, K.D. (2007). Local sociotechnical system development in the NHS National Programme for information technology. *Journal of Information Technology* **22**(3), 257–264.
- Eason, K.D., Dent, M., Waterson, P., Tutt, D., Hurd, P. and Thornett, A. (2012). Getting the benefit from electronic patient information that crosses organisational boundaries. Final Report. London, U.K.: NIHR Service Delivery and Organisation Programme.
- Eason, K.D., Harker, S.D. and Olphert, C.W. (1996). Representing socio-technical systems options in the development of new forms of work organization. *European Journal of Work and Organizational Psychology* **5**, 399–420.
- Eason, K.D. and Sell, R.J. (1981). Case studies in job design for information processing tasks. In Corlett, E.N. and Richardson, J. (eds.), *Stress, Work Design and Productivity*. Chichester, U.K.: John Wiley, pp. 195–208.
- Eason, K.D., Yu, L. and Harker, S.D.P. (2000). The use and usefulness of functions in electronic journals: The experience of the super journal project. *Program* **34**, 1–34.
- Emery, F.E. (1980). Designing socio-technical systems for ‘Greenfield’ sites. *Journal of Occupational Behaviour* **1**, 19–27.
- Galer, M., Harker, S.D. and Ziegler, J. (1992) (eds.). *Methods and Tools in User-Centred Design for Information Technology*. North-Holland, the Netherlands: Elsevier.
- Haeckel, S.H. (1999). *Adaptive Enterprise*. Boston, MA: Harvard Business School Press.
- Hall, G.E. and Hord, S.M. (2006). *Implementing Change: Patterns, Principles and Potholes*. Boston, MA: Pearson/Allyn & Bacon.
- Hoeksma, J. (2008). Royal free says £7.2 m deficit due to Cerner E-Health insider, 5 November 2008. [www.ehi.co.uk/news/ehi/4307](http://www.ehi.co.uk/news/ehi/4307), accessed 21 September 2012.
- House of Commons Transport Select Committee. (2008). *The Opening of Terminal 5*. London, U.K.: The Stationary Office.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Katz, D. and Kahn, R.L. (1966). *The Social Psychology of Organizations*. London, U.K.: Wiley.
- King, A. and Crewe, I. (2013). *The Blunders of Our Governments*. London, U.K.: Oneworld.
- Klein, L. and Eason, K.D. (1991). *Putting Social Science to Work*. Cambridge, U.K.: Cambridge University Press.
- Koukoulaki, T. (2014). The impact of lean production on musculoskeletal and psychosocial risks: An examination of sociotechnical trends over 20 years. *Applied Ergonomics* **45**(2A), 198–212.
- Latour, B. (1986). The powers of association. In Law, J. (ed.), *Power, Action, and Belief: A New Sociology of Knowledge*. London, U.K.: Routledge & Kegan Paul.

- Lewin, K. (1947). Frontiers of group dynamics. *Human Relations* **1**, 5–41.
- Luecke, R. (2003). *Managing Change and Transition*. Boston, MA: Harvard Business School Press.
- Mumford, E. (1983). Designing human systems-The ETHICS method. *Manchester Business School*, Manchester, U.K.
- Mumford, E. (2000). Technology and freedom: A socio-technical approach. In Coates, E., Willis, D. and Lloyd-Jones, R. (eds.), *The New Socio-Tech: Graffiti on the Long Wall*. London, U.K.: Springer-Verlag, pp. 29–38.
- NAO. (1999). The passport delays of summer 1999, Report of HC session 1988–1999, HC 812. London, U.K.: National Audit Office, The Stationery Office.
- NAO. (2008). The National Programme for IT in the NHS: Progress since 2006. London, U.K.: National Audit Office, The Stationery Office, 16 May 2008.
- NAO. (2010). FiReControl project. London, U.K.: National Audit Office, The Stationery Office), 10 February 2010.
- NAO. (2011). The National Programme for IT in the NHS: An update on the delivery of detailed care records systems. London, U.K.: National Audit Office, The Stationery Office, 18 May 2011.
- Orlikowski, W. (2000). Using technology and constituting structures: A practice lens for studying technology in organisations. *Organisation Science* **11**(4), 404–428.
- Oudshoorn, N. and Pinch, T. (eds.) (2003). *How Users Matter: The Co-Construction of Users and Technology*. London, U.K.: MIT Press.
- Page, D., Williams, P. and Boyd, D. (1993). Report of the inquiry into the London ambulance service. Report commissioned by South Thames Regional Health Authority, London, U.K.
- Parker, S.K. and Wall, T.D. (1998). *Job and Work Design*. London, U.K.: Sage.
- Pinch, T. and Bijker, W. (1987). The social construction of facts and artefacts. In Bijker, W., Hughes, T. and Pinch, T. (eds.), *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press, pp. 17–50.
- Revens, R.W. (1982). *The Origins and Growth of Action Learning*. Bromley, U.K.: Chartwell Bratt.
- Ritter, T. (2008). ASSIST says NHS one-size-fits-all doesn't work, *Computer Weekly*, 7 October 2008. <http://www.computerweekly.com/blogs/public-sector/2008/10/assist-says-nhs-onesizefitsall.html#more>, accessed 24 November 2014.
- Rothlisberger, F.J. and Dickson, W.J. (1939). *Management and the Worker*. Cambridge, MA: Harvard University Press.
- Senge, P. (1990). *The Fifth Discipline*. New York: Doubleday.
- Tapscott, D. (1982). *Office Automation: A User Driven Method*. New York: Plenum.
- The Standish Group. (2014). Chaos report. <http://www.projectsmart.co.uk/docs/chaos-report.pdf>, accessed 24 November 2014.
- Wastell, D.G. and Cooper, C.L. (1996). Stress and technological innovation: A comparative study of design practices and implementation strategies. *European Journal of Work and Organizational Psychology* **5**, 377–397.
- Wilson, J. (1994). Devolving ergonomics: The key to ergonomics management programmes. *Ergonomics* **37**, 579–594.



---

# 33 Methods in Systems Ergonomics

*Carys E. Siemieniuch and Murray A. Sinclair*

## CONTENTS

|                                                                                            |     |
|--------------------------------------------------------------------------------------------|-----|
| Introduction.....                                                                          | 856 |
| Definition of Important Terms Used in the Chapter .....                                    | 856 |
| System .....                                                                               | 856 |
| Systems of Systems.....                                                                    | 858 |
| Systems Ergonomics .....                                                                   | 859 |
| Systems Engineering.....                                                                   | 860 |
| Systems Thinking.....                                                                      | 860 |
| Significance of Systems in Society .....                                                   | 861 |
| Roles for Humans in Systems .....                                                          | 861 |
| Benefits of Including Humans in Systems .....                                              | 862 |
| ‘Costs’ of Human Inclusion in Systems.....                                                 | 862 |
| When to Use Systems Ergonomics Methods.....                                                | 863 |
| System Life Cycle.....                                                                     | 863 |
| System Development Process .....                                                           | 864 |
| Nature of Systems Interventions.....                                                       | 865 |
| Classes of System Intervention .....                                                       | 865 |
| Standard Process Instructions (SPIs) and Standard Operating Procedures (SOPs).....         | 866 |
| Qualitative and Quantitative Methods.....                                                  | 867 |
| Experimental Paradigm Is Not Always the Right Approach for a Systems Ergonomics Issue..... | 867 |
| Standards in Systems Ergonomics .....                                                      | 868 |
| Classes of Methods in Systems Ergonomics.....                                              | 868 |
| Methods Handbooks.....                                                                     | 868 |
| Sources of Information for Systems Ergonomists .....                                       | 871 |
| Methods for Collecting Data.....                                                           | 871 |
| Systems Design: Methods from Systems Engineering and Other Disciplines .....               | 873 |
| Requirements Engineering .....                                                             | 874 |
| Network Models .....                                                                       | 875 |
| Discrete Event Simulation.....                                                             | 875 |
| Agent-Based Modelling .....                                                                | 875 |
| System Dynamics.....                                                                       | 876 |
| Human Cognitive Performance Modelling .....                                                | 876 |
| Process Analysis.....                                                                      | 876 |
| Physiological Assessment Methods.....                                                      | 877 |
| Other Useful Techniques .....                                                              | 877 |
| Conclusions.....                                                                           | 877 |
| References.....                                                                            | 877 |

## INTRODUCTION

As the title says, this is a chapter mainly about methods. But all methods have their strengths and weaknesses, and work best only for a range of situations. Recognising this, the chapter is written to describe the context in which the methods would be used, and then outlines a number of methods likely to be of general use for systems ergonomics studies.

The utilisation of these methods is also discussed in other chapters in this book; in particular, Chapter 9 (Systems Analysis and Requirements Specification), Chapter 10 (Involving People in Design Research), Chapter 28 (Socio-Technical Design of Work Systems) and Chapter 32 (Ergonomics Interventions in the Implementation of New Technical Systems).

The reader is urged to read these chapters in conjunction with this chapter before applying the methods discussed later.

The structure of the chapter is as follows. The next section provides a set of definitions that the authors have found to be appropriate, covering systems, system of systems (SoS), systems ergonomics, systems engineering and system thinking. This is followed by an introduction to the importance and role of humans in systems which in turn is followed by a guide on when to use systems ergonomics methods. The chapter then moves on to a consideration of the types of systems interventions that may be required and how systems ergonomics methods fit into these. Systems interventions must incorporate standards, of which some important ones are listed in the next section. Finally, a range of systems ergonomics methods are introduced with references for further reading, and a brief conclusion ends the chapter.

## DEFINITION OF IMPORTANT TERMS USED IN THE CHAPTER

The domain of systems, in its widest sense, is of profound interest to a multitude of different disciplines. This has the unfortunate corollary that the fundamental concepts can be defined with different emphases, depending on the discipline. For example, as Lars Skyttner has explained,

A system is not something presented to an observer; it is something to be recognised. Most often the word does not refer to existing things in the real world but rather to a way of organising our thoughts about the real world. The constructivist view of reality states that systems do not exist in the real world independent of the human mind;... the fictionalist view takes a further step and states that the systemic concept can be well suited to its purpose even if we know that it is incorrect or full of contradictions in a specific situation. (Skyttner 1996)

For clarity, this chapter is based on the following definitions, all of which are widely accepted among the many disciplines. For each concept, we provide a definition and give some commentary on its usage.

### SYSTEM

A system might generally be described as an organised set of elements or subsystems interconnected by relationships to form a whole – a system has a purpose but this purpose may depend on the perception of the observer.

Although the title of this chapter includes the term systems ergonomics, we have chosen to use the systems definition from the International Council on Systems Engineering (INCOSE), the reason being that much of ergonomics/human factors (E/HF) thinking is eventually implemented by engineers. Accordingly, the definition for this chapter is

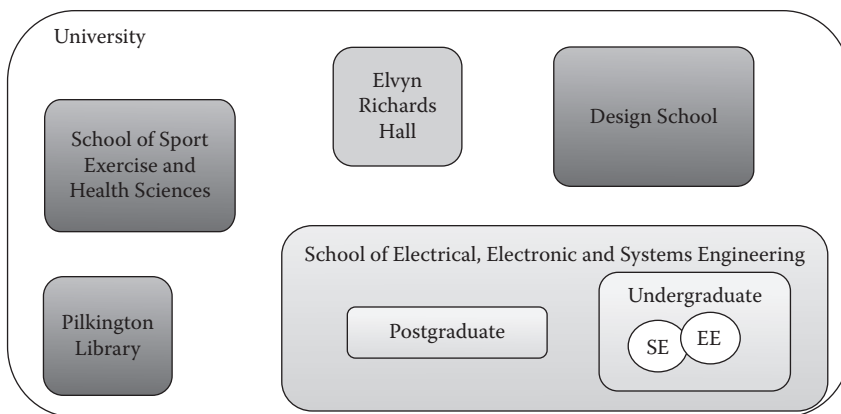
An integrated set of elements to accomplish a defined objective. These include hardware, software, firmware, people, information, techniques, facilities, services and other support elements. (Haskins 2010)



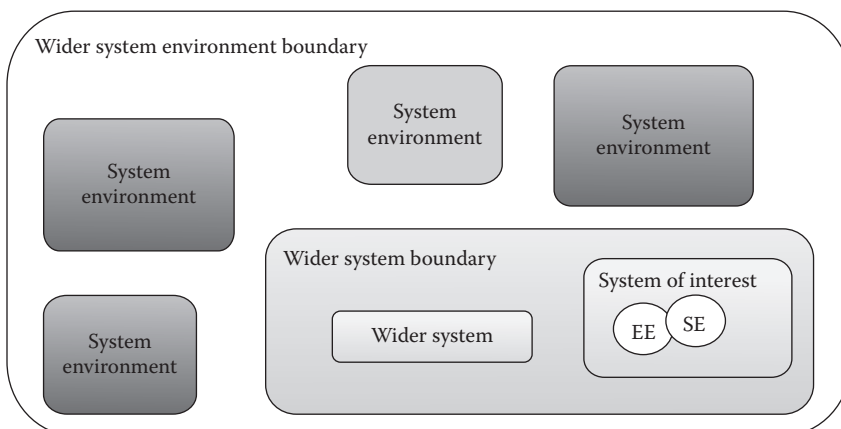
While this definition is a widely used definition of a system, it does not specifically mention the notion of a system boundary demarcating what is considered to be in the system of interest, and everything else that performs is a part of the immediate and wider environment of the system. As Skyttner has implied, the boundary is a creation of the observer or system developer and is therefore movable, and may be different for different people.

For example, Figure 33.1 shows a university as the overall system within which there are several subsystems (schools, library or halls of residence). Figure 33.2 shows that if the undergraduate teaching system within the School of Electrical, Electronic and Systems Engineering is the main system of interest, then immediate and wider system boundaries can be put in place to aid the focus of the developer. Systems in the wider system environment cannot be ignored, and the developer must understand the interactions between these and the system of interest and particularly the impact these interactions may have on any functionality within the immediate system of interest. If the system of interest is the library system, then the nature of interactions of interest between this system and other systems in the system environment would be different.

A corollary to this definition of a system is that although we may be able to identify a system, we may not have a good understanding of its behaviour. It is usually wise to assume that this is the case for any system of interest, and spend time on improving the understanding of the system before tinkering with it.



**FIGURE 33.1** Overview of a sample set of systems that make up a university system.



**FIGURE 33.2** Representation of the system of interest and its boundaries.

Nor does the definition mention the word ‘control’. Typically, it is the control that is built into the system that enables its behaviour to change to continue meeting the system’s objectives despite changing environmental conditions or changes in the objectives themselves. The control of the system is usually expressed hierarchically over the system (i.e. ‘sits on top of’) and its component subsystems.

Finally, the definition does not discuss ownership of the system. Usually the system has a single owner, who is responsible for its performance (on time, in full, to the required quality and at specified cost) and for its sustenance over its life cycle, and who decides when and to what extent changes will be made and at what cost.

But what the definition does imply is that we need to consider all these elements interacting together to achieve a given purpose; that this purpose is different (usually larger) compared to any of the purposes of the individual elements or subsystems; and that therefore we have to adopt a holistic viewpoint. The traditional approach of decomposing the system into its elements and then looking at each may be necessary, but it is insufficient. This has implications for the methods outlined later.

As a final comment in this section, the reader may find it useful to refer to the special issue on *Systems Ergonomics/Human Factors (Applied Ergonomics, Volume 45, Issue 1)* for contemporary examples of systems ergonomics thinking.

## SYSTEMS OF SYSTEMS

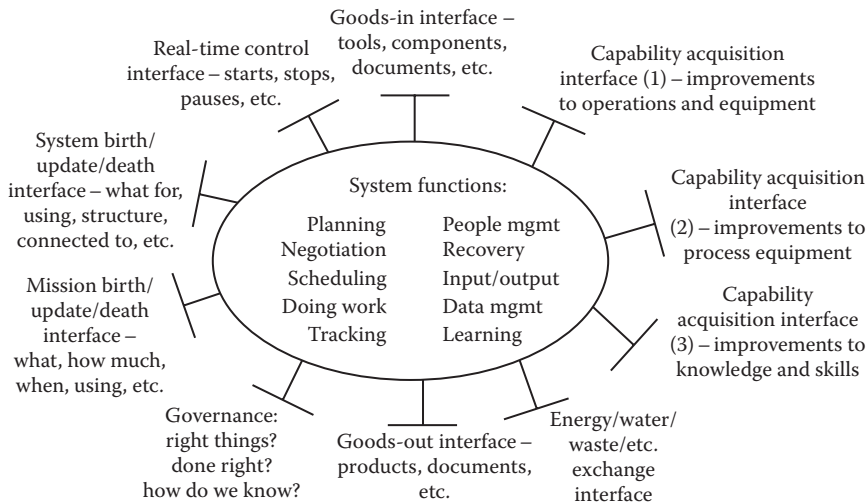
What distinguishes an SoS from a system is the ownership. In an SoS, each system has a different owner, responsible for the performance and upkeep of that system alone. Consequently, control of the whole SoS in principle can only be exerted at the interfaces between the systems. The classic SoS definition that distinguishes SoS from systems is as follows (Maier 1998):

- The elements of the system are themselves sufficiently complex to be considered systems.
- Operating together the systems produce functions and fulfil purposes not produced or fulfilled by the elements alone.
- The elements possess operational independence. Each element fulfils useful purposes whether or not connected to the assemblage. If disconnected, the element continues to fulfil useful purposes.
- The elements possess managerial independence. Each element is managed, at least in part, for its own purposes rather than the purposes of the collective.

There are a few implications that follow from this definition: firstly, at the SoS level, individual systems may become ‘black boxes’ (i.e. the system exists, and you can use it, but you are not allowed to see inside it), for reasons such as security or commercial confidentiality. In this case, it is only at the interfaces where transparency occurs, open analysis of performance happens and control can be exerted, and then only by negotiation and by reference to the terms expressed in any contracts. Secondly, managerial independence implies that each system will be optimised for its own benefit, not necessarily for the benefit of the whole SoS. Thirdly, most of the systems on which society depends fall under the SoS banner; health, energy, food, policing and justice, and transport are all examples. These cannot be allowed to fail, and there is no termination date for these SoS; effectively they are immortal, with all the problems of continuity represented by this.

Figure 33.3 is relevant here. Within the oval is in the domain of the component system, and the T-bars outside the oval denote the generic interfaces that the system requires in its interactions with the rest of the SoS and with the environment. From the SoS perspective, what happened within the oval is ‘hidden’; it is the interfaces which are important, with the implication that each of them must connect to some system outside, if the interface is to be functional and the SoS is to reach its goals. It is for this reason that defining a boundary to the SoS can become difficult.

In practice, although there is a clear distinction between systems and SoS in the given definitions, the distinction is much less clear when one looks at real-world systems. When the system is large



**FIGURE 33.3** Representation of a system and its generic interfaces with its environment.

(e.g. an organisation with thousands of employees), with lots of parallelism in the operations of the subsystems, and with authority over the subsystems organised into vertical silos, the system can look and behave more like an SoS. Likewise, when there is a dominant partner in the SoS, because of its size and buying power, for example, this partner may be able to dictate to all the other participants, and the SoS looks much more like a system (Dahmann and Baldwin 2008). These issues and more are discussed in Jamshidi (2009a,b) and Barot et al. (2012).

## SYSTEMS ERGONOMICS

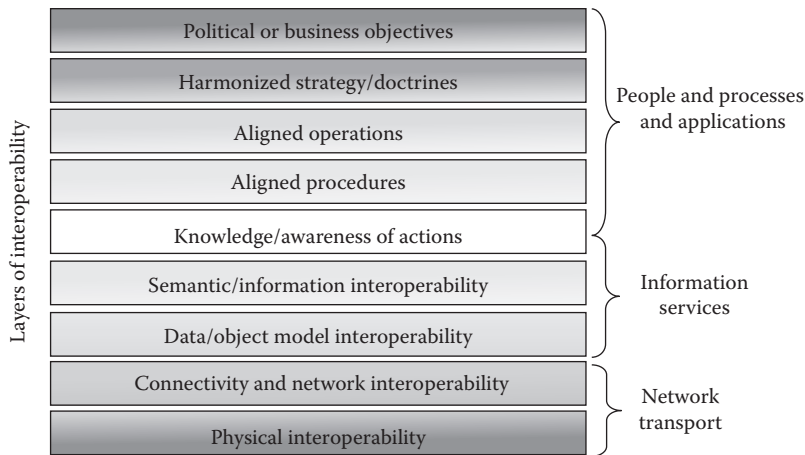
Systems ergonomics falls within the remit of systems engineering (see later section); however, its focus is very much on the roles, skills and knowledge, awareness and communications, the distribution of responsibilities and authorities over resources and support of the humans that function within the system to achieve the system's goals, and to the people, organisations and society outside the system boundary that may be affected in some way by the operation of the system. Another way to express this is to rephrase the definition of a system into the following:

A system is a group of people working to a common purpose, supported by a network of peripherals.  
(S. Deal, CEO DEAL Corp., USA)

This definition emphasises the fundamental role of people in systems, and this focus is what distinguishes systems ergonomics as a component of systems engineering.

Another way to distinguish systems ergonomics is to make use of the NCOIC Interoperability Framework (NIF), which was implemented as a development framework which helps system architects and system engineers to ensure interoperability throughout the life cycle of systems and SoS programmes, beginning with requirements (NCOIC 2011). This is shown in Figure 33.4. Systems ergonomics has a focus mainly (but not solely) on the top five levels of this diagram, whereas systems engineering deals with all levels, but in practice mainly from a technical perspective. This leads to the following definition for system ergonomics used by the authors for this chapter:

Systems ergonomics is aimed at the optimisation of socio-technical systems, from direct interactions between humans and system technology through to the organisation and support provision for humans to attain the purposes of the system.



**FIGURE 33.4** The NIF layers as defined in 2008. Systems ergonomics concentrates on the top five layers, whereas systems engineering considers all of them, but with an emphasis on the technology. (From NCOIC, *NCOIC Interoperability Framework 2.1*, Network Centric Operations Industry Consortium, Washington, DC, 2011, <https://www.ncoic.org/technology/deliverables/nif/>.)

## SYSTEMS ENGINEERING

INCOSE has defined systems engineering as follows:

Systems engineering is an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, then proceeding with design synthesis and system validation while considering the complete problem: Operations, Cost and Schedule, Performance, Manufacturing, Training and Support, Test, and Disposal. Systems engineering integrates all the disciplines and specialty groups into a team effort forming a structured development process that proceeds from concept to production to operation. Systems engineering considers both the business and the technical needs of all customers with the goal of providing a quality product that meets the user needs.

This definition has no formal emphasis on the human and organisational components of the system (though good systems engineers always pay attention to these), and it is here that systems ergonomics makes its contribution.

Systems engineering is a complex engineering discipline, but essentially it is about making systems work the way they are supposed to work with no nasty surprises – it is about engineering the system as a whole over the whole life cycle of a system.

## SYSTEMS THINKING

Systems thinking, as the name suggests, is a way of envisioning the world. Its definition is as follows (Senge 1990):

Systems thinking is a discipline for seeing wholes. It is a framework for seeing interrelationships rather than things, for seeing patterns of change rather than static ‘snapshots’.

The principal aspect of systems thinking is that the world is seen holistically, with an emphasis on feedback loops, delays and patterns of dynamic behaviour. This is in contrast to the orthodox approach of decomposition of the objects of attention into parts, and then analysis of the parts,

in order to understand what is going on. Systems thinking has led to a number of principles and laws of importance when selecting methods to explore, build and operate systems (Conant 1981, Skyttner 1996). Among the most important are the following:

- *Interdependence of objects and their attributes*: Independent elements can never constitute a system.
- *Hierarchy*: Complex wholes are made up of smaller subsystems.
- *Complementarity law*: Any two different perspectives (or models) about a system will reveal truths about the system that are neither entirely independent nor entirely compatible.
- *Law of requisite variety (Ashby's law)*: Control can be obtained only if (a) the variety of the controller is at least as great as the variety of the situation to be controlled, and (b) there is sufficient channel capacity in the system.
- *Darkness principle*: No system can be known completely.
- *Equifinality*: Systems should have alternative ways of attaining the same objectives (convergence).
- *Multifinality*: It should be possible to attain alternative objectives from the same inputs (divergence).

One other view on systems thinking is worth considering:

Systems Thinking is the art and science of making reliable inferences about behaviour by developing an increasingly deep understanding of underlying structure.

**B. Richmond, CEO, High Performance Systems**

## SIGNIFICANCE OF SYSTEMS IN SOCIETY

Arguably, systems have been in existence from when our human ancestors could reason about their environment, and change it. However, the systems approach became formalised in the 1940s and has developed ever since (Ashby 1956, Bertalanffy 1968). Systems themselves have become pervasive in society, greatly aided by the developments in information technologies and communications, to the point where society itself depends on their continuous, reliable functioning. The delivery of health care, energy, the supply of provisions, education and government itself all now depend on systems and, increasingly, on their interoperation in networks of systems, redefined as SoS here. In turn, the effective and efficient deployment of these systems and their successors depends on the knowledge, skills and gleaned experience of people applying the disciplines defined earlier using the methods described in this and other chapters in this book.

Systems, and by extension SoS, are likely to be of ever-increasing importance, as the Sulston report has indicated (Sulston 2012); by 2050, it is hoped that the global human population will have stabilised at 10 billion people, although with attendant problems for humans due to climate change and sustainability issues (McMichael and Dear 2010, Sherwood and Huber 2010, Allwood et al. 2011). It is likely that a sustainable equilibrium will only be accomplished by an extensive complex of systems and SoS, delivered by appropriately trained professionals.

## ROLES FOR HUMANS IN SYSTEMS

The benefits and costs are discussed later, relative to systems. This is because it is a purpose of systems ergonomics to make sure that the unique benefits of human involvement in systems are delivered while ensuring that the 'costs' that come with these unique benefits are properly addressed.

## **BENEFITS OF INCLUDING HUMANS IN SYSTEMS**

The benefits of including humans in the organisation, operation and support of systems are well known. Perhaps the best examples happened in the 1980s, when several manufacturers tested the notion of ‘lights-out’ factories (Thomas 1994). These factories, where humans were replaced on the night shift by automation, were demonstrated to be non-cost-effective, usually because some fault that the automation could not address would cause the factory to come to a halt at some point in the night. It is this ability to deal with the unexpected that makes humans so valuable in systems (Thomas 1994).

Decomposing this, the benefits accruing from the involvement of humans in the whole life cycle of systems can be stated as the provision of thinking, decision-making, responsibility-accepting, behaviour-adapting, problem-identifying, innovating, information-processing, equifinality-capable and multifinality-capable, complexity-absorbing, goal-seeking, learning and ethical entities. Of these, perhaps the most significant benefit is ‘responsibility accepting’ because under the legal systems of the majority of countries it is only humans who can accept responsibility. This is of some importance; as robots and other cyber-physical systems become a part of SoS, there must still be responsibility for their actions, especially where action may take place at a distance, across legal boundaries (Asaro 2011, 2012). Further discussion of the issues involved when humans and (semi) autonomous systems are cooperating can be found in Hollnagel and Woods (2005) and Woods and Hollnagel (2006), within the framework of ‘joint cognitive systems’.

Secondly, it is the incapacities of robots and other semi-autonomous, cyber-physical systems for ethical behaviour that emphasise the importance of humans within systems and SoS. Given that a system will have several people within it, and will interoperate with external systems that also contain people in positions of responsibility, then it is ethical behaviour that will enable these systems to work most effectively in non-stable environments where collaborative decision-making is required. This is because ethical behaviour is a foundation on which trust can be built, and without which it is difficult to negotiate effective solutions to discovered problems.

Thirdly, it is the human capacity to learn. A fundamental rule was enunciated some 2500 years ago: ‘You cannot step in the same river twice’ (Heraclitus 535–475 BCE, Davenport 1979), more commonly known as ‘nothing stays the same’. Given the truth of this, then learning is a necessary requirement for any long-lasting system. The advantage of humans is that they are capable of triple-loop learning, and despite the advances of information technology, this is still a humans-only capability. Single-loop learning is equivalent to learning how to execute the steps of a process; second-loop learning is concerned with learning how to adapt the process to make it more effective, as in the mantra ‘faster, better, cheaper’. Third-loop learning comes into play when it is realised that the current process has reached its limit and a different way of doing things must be deployed. It is in this latter paradigm-shifting knowledge where humans excel.

## **‘COSTS’ OF HUMAN INCLUSION IN SYSTEMS**

While humans deliver the essential systems skills outlined earlier, they, too, need support as does every system. The ‘costs’ that come with the capabilities are the need for support (at task, organisation and system levels), the occurrence of ‘human error’ (Perrow 1999, Weick et al. 1999, Kirwan 2005, Weick and Sutcliffe 2007, Reason 2008) though this is often unintentionally ‘designed into’ systems as a result of complexity, changing circumstances and, as is usually seen in hindsight; insufficient application of systems engineering, slowness of cognition in fast-changing conditions, the effects of fatigue, limited capability to cope with complexity (Siemieniuch and Sinclair 2002, Allen et al. 2005, Richardson 2005) and the never-ending need for knowledge capture and renewal because people come and go, and nothing stays the same (Sinclair 2005). However, it is a specific purpose for systems ergonomics to address these ‘costs’ in order that the benefits, many of them unique and vital to the system, can be realised; in fact, this is the purpose of this whole book.



## WHEN TO USE SYSTEMS ERGONOMICS METHODS

### SYSTEM LIFE CYCLE

All systems have a life cycle; a generic version of this is the CADMID life cycle, from the UK Ministry of Defence. CADMID is an acronym for Concept, Assessment, Demonstration, Manufacture, In-service and Disposal:

*Concept:* The intended purpose and utilisation of the system and its relationships with other new and existing systems are determined. It may not be a novel concept; more often it will be a replacement concept, superseding and extending a capability that already exists. An automobile company's products provide an example of this.

*Assessment:* Continuous evolution and evaluation of the concept through experiments, simulations and prototypes. In the military domain, assessment covers the eight 'lines of development' (LoD), namely training, equipment, personnel, information, doctrine (policy, in the civilian world), organisation, infrastructure and logistics. Assessment includes both meeting the requirements for each of the LoDs raised by the concept and the interactions between the LoDs that exist in a delivered system.

*Demonstration:* This is the final prototyping phase, where a working example of the concept is evaluated to show that the capability that the concept embraces is actually being delivered.

*Manufacture:* The stages above determine the costs of the project; this stage is where the bulk of the expenditure happens. Because several years may have passed by the time this stage starts, and more will pass before it completes, it is possible that some redesign will be necessary as the world around the project changes. It is for this reason, and because of in-service upgrades, that it has been said that 'there is no navy in the world that has two identical submarines'. The significance of this is illustrated in the following example, condensed from a document published by Sun Microsystems™.

In the 1980s, a nuclear submarine might have had 8 million parts and each submarine was different. Information for these was held in about 200,000 US Navy manuals, for the next 40 years. Revisions to these manuals amount to several million changes per year. Because of the inevitable delays and dislocations of administrative processes, the gap between the manuals and reality steadily widened. Consequently, it was estimated that 25% of the documentation was inaccurate or out of date. As a result, 5%–8% of all accidental military deaths could be traceable to errors in publications. Furthermore, between 1977 and 1986, 47% of major USAF accidents were deemed to have been caused by inaccurate technical information (Sun Microsystems 1991).

Subsequent improvements in logistics management such as CALS – Continuous Acquisition and Lifecycle Support (DTI 1992) – have addressed this problem, but it is unlikely to have gone away completely.

*In-service:* This stage covers the deployment, operation and maintenance of the capability that has been developed from the original concept. Inevitably, because time has passed, because design seldom is able to encompass the full complexities of the real world and because the world will have changed anyway, upgrades and fixes will be necessary for the system. These happen through the CAD stages given earlier and are steadily incorporated into the delivered systems. There is the classic example of the B-52 bomber, conceived as an intercontinental strategic nuclear bomber in the 1950s, and still operational today, but adapted to be a high-altitude, close air support capability for ground troops, with an appropriately reconfigured support network. It is also noteworthy that given the near 60 years' service lifetime of this aircraft type, there will be nobody actively involved in the support of its capability who was there at the beginning, and much of the accumulated knowledge may be held in storage systems no longer accessible with today's technology – a problem in the conservation of knowledge and wisdom.

*Disposal:* This covers the removal from active service of the system and the capability that it embodies; in other words, ending all the functions and discontinuing and disentangling all the interfaces shown in Figure 33.3. This may mean just re-assigning and retraining the personnel involved, and placing the hardware in storage to gather dust; it may also mean removing components dangerous to health and safety and rendering them inert using processes that may take many years. There are documentation issues, too; the current disposal exercise might happen because a new, improved system is being installed, and all the documentation for those systems that will interoperate with this new system will have to be updated, including all the components that have been changed in these systems over the years.

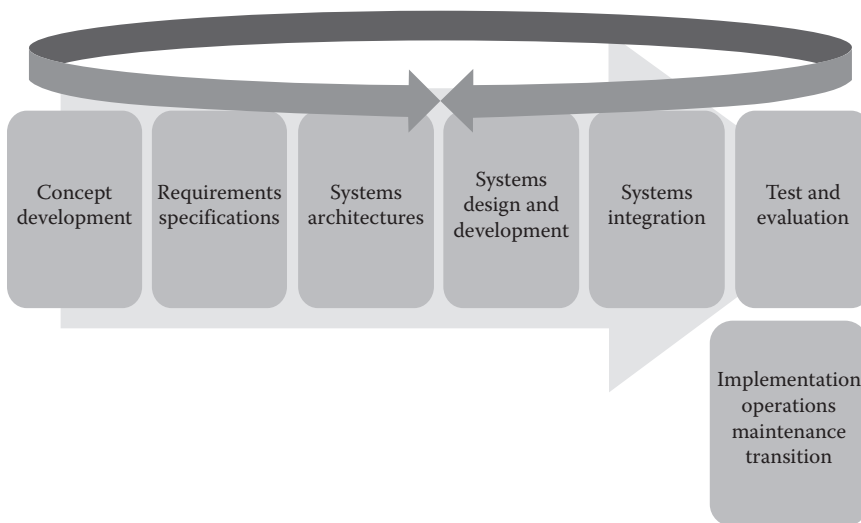
## SYSTEM DEVELOPMENT PROCESS

Figure 33.5 provides a representation of a generic system development process, representing the means by which the CADMID life cycle is delivered. Systems ergonomists have contributions to make in all the stages, but perhaps the key areas are concept development, requirements engineering, system design and development, and test and evaluation.

*Concept development:* Concerned with transforming a user's expression of an operational need into a well-defined concept of operations, a high-level conceptual definition and a set of initial operational requirements.

*Requirements engineering:* The key activities are to elicit detailed system requirements from the users and other stakeholders, to analyse and refine the requirements and to develop plans and a process for managing the requirements throughout the rest of the system life cycle. Complex systems always demonstrate a degree of instability and uncertainty with requirements, so methods to accommodate this are needed.

*System architecture:* Once the requirements are expressed, a system architecture can be described which will be the foundation for further development, integration, testing, operation, interfacing and improvement of the system as time goes on. Various types of system architectures (e.g. service-oriented architecture) and architectural frameworks, such as the U.S. Department of Defence Architecture Framework (DODAF), the British Ministry



**FIGURE 33.5** System design life cycle: the background arrow implies progress through the stages of the system development process. However, there will be considerable iteration through the stages as shown by the circular arrowheads at top, as extra requirements and constraints appear.

of Defence Architecture Framework (MODAF) and TRAK, exist. A formal process for developing architectures is required.

*System design and development:* At this point in the system life cycle, the following has been developed: a complete and comprehensive description of what and how the system is expected to perform (requirements specification), and an architectural representation to guide the actual design and development of the hardware, software and interfaces. Therefore, this stage is about the following:

- Developing system-level technical requirements
- Assessing design's ability to meet the system requirements
- Development of a top-level system design
- Sub-system development

*Systems integration:* This stage is about integrating the system's components and its interfaces with other systems into an operational whole. It will be necessary to identify and assess integration and interoperability challenges, develop and evaluate integration and interoperability solution strategies, assess integration testing approaches and interface management.

*Test and evaluation:* Often referred to as verification (demonstrate that the system, as made, is 'right', i.e. fulfils the specified requirements) and validation (demonstrate that the 'right' system has been created, i.e. is fit for purpose), abbreviated to V&V. V&V is a whole life cycle process and must be applied at each stage in the systems development process. It has three principal objectives:

- The discovery of defects in a system
- The assessment of whether or not the system is fit for purpose in an operational situation
- Reassurance of customers and users that the system will not cause problems

*Implementation, operations, maintenance and transition:* The aim here is to ensure a successful transition of the system into the field and hence plans and procedures must be developed for operations and maintenance.

Note that since the technological underpinnings of a system are constantly changing, product improvements – including the insertion of new technologies – must be planned.

## NATURE OF SYSTEMS INTERVENTIONS

It is a truism that the need for intervention in systems never goes away, and some of the reasons have been mentioned earlier. The extent of the intervention may vary, and so might the reason for the intervention. It is the latter that we discuss next.

### CLASSES OF SYSTEM INTERVENTION

There are three generic classes of system intervention, listed later. It should be noted that although these are three stand-alone interventions, it is often the case that an intervention will include more than one of these classes:

- *Incrementally improving/upgrading systems:* The B-52 example is an illustration of this. There is a design principle to be remembered for this class, which will save both time and money, first enunciated by Rasmussen (Rasmussen and Goodstein 1986):

We should regard the operator as the system designer's on-site representative.

- Chapter 14 in this book has more to say on this. It is this operational knowledge about the system's problems, workarounds, (un)reliability, resilience and needed fixes that are the basis for the next iteration of upgrades to the system.

- *Designing systems ab initio*: This is the ‘blank sheet of paper’ scenario, but it is usually a mistake to adopt this approach literally. This class of intervention usually occurs as a part of the system replacement scenario, where an old system is being retired and a new, more capable system is going to be inserted. As a general rule, all the functions and unexpected events that the old systems had to accommodate will still have to be included in the new system, together with any new functionality. If the old system was in place for more than a year or two, it is likely that the users, operators and other stakeholders of that system will have picked up useful experience and skills in making the system work, perhaps for unexpected uses, and fixing problems. This experiential knowledge is a very valuable source of understanding for the design of new systems, and this has been encoded into a design principle:

There is always a precursor system: Find it, and mine it for knowledge, lest you have to re-learn this painfully, later. (A. Gardner, 1990, private communication)

- *Accommodating legacy systems that cannot/should not be changed*: This is a particular issue for bespoke systems, designed at some time in the (distant) past for a particular purpose for a particular user. It is likely that no one from the design team is still available, and that none of the original documentation is still complete and correct. The latter is certainly likely for IT systems where various coders have patched someone else’s code without understanding all the ramifications of the application. It takes little imagination to foresee the legacy problems that could occur, given the following quotation:

In one case, we observed an outsourced application with 120 COTS products, 46% of which were delivered in a vendor-unsupported state. (Yang et al. 2005)

- The issues of legacy systems are widespread, and as more and more systems are inter-linked, so the consequences of their failure can ramify widely. In hardware terms, this is shown quite well by some of the publicised electrical grid failures that have affected millions of people (Andersson et al. 2005, Romero 2012).

## STANDARD PROCESS INSTRUCTIONS (SPIs) AND STANDARD OPERATING PROCEDURES (SOPs)

A good metaphor to distinguish between these is the fully equipped kitchen. For the chef in this kitchen, standard process instructions (SPIs) are the recipes that result in the meals that the customers have ordered. Standard operating procedures (SOPs) are the instructions for safely and correctly operating the equipment in the kitchen that enables the recipes to be created and delivered for the customers.

In this chapter, we concentrate on the SOPs for systems ergonomics. In other words, we describe the methods that could be used, but not the processes in which they would be used. For instance, it is for this reason that the Soft Systems Methodology (Checkland 1995), the Rigorous Soft Methodology (Hitchins 2003) and the standard INCOSE VEE diagram (Haskins 2010) are not included in this chapter. Nor have we included architectural frameworks, such as the Zachman Framework (Sowa and Zachman 1992, Azani 2008), DODAF (Ohren 2004) and MODAF (Caseley et al. 2007, Bruseberg and Fletcher 2008), on the grounds that they are not methods, although they organise systems thinking.

Staying with the kitchen metaphor, it will be understood that each piece of equipment in the kitchen has a limited range of uses. This too will be included in the discussion about methods, but it should be noted that generally methods are much more flexible than hardware, so some care will be necessary when reading the comments on method applications.

## QUALITATIVE AND QUANTITATIVE METHODS

Once humans are included in systems, as inevitably they will be, one must include consideration of social as well as technical aspects because of the essential roles of humans as agents of responsibility, authority, change management, control or operation; necessarily, then, one must consider the whole human agent, embedded within an organisational structure (which itself is embedded within a social order), capable of autonomous behaviour and equipped with a personal history and set of attributes, knowledge, experience and skills. Essentially, this is a world of relationships, ethics and perceived duties, in which physics and engineering principles and laws play little part, and constancy is an abstract concept. Furthermore, accessing this inner world is not easy; typically we are reduced to measuring human performance outputs and inferring what is going on inside; we may be able to make physiological measurements of the human(s) and make use of physiological models to assess internal workings; we may analyse human communications and control actions within the system, and we can ask people to tell us what they are experiencing directly, through the use of questionnaires and scaling techniques of one sort or another – Chapter 4 in this book is relevant here.

It is certainly possible to turn all this information into numbers, but whenever one does this using the methods described later, one should bear in mind the sceptic's mantra:

What does this number mean, and for how long does it mean it?

For ergonomics/ human factors (E/HF) professionals, while the structures of the system are of interest, it is the dynamics of the system – its behaviour over time – that is critical.

While E/HF engineers are usually happy to work with qualitative data, many other classes of engineers are not, because it means leaving the world of physics in which most of their expertise lies. For this reason, it is incumbent upon ergonomists and human factors professionals to translate their findings and numbers into a format and a dialogue that has meaning for the relevant engineers; after all, it is the latter who play a major part in instantiating the work of the former into working systems that meet the needs of society.

## EXPERIMENTAL PARADIGM IS NOT ALWAYS THE RIGHT APPROACH FOR A SYSTEMS ERGONOMICS ISSUE

Because of the standard training that human factors personnel receive, there is a strong tendency for us to turn to experiments as a first resource to discover information that is needed to solve a systems ergonomics problem. However, this urge needs to be tempered, both because of the essentially holistic nature of systems and especially in the light of what is now known about complexity in systems (Chapanis 1967, Rasmussen 1994, Allen 1998, Arrow et al. 2000, Edmonds 2000, Cilliers 2005, Richardson 2005).

For example, consider the following definition of complexity (Allen 1998):

Complexity is really just reality without the simplifying assumptions that we make in order to understand it.

The dangers of simplification are self-evident, especially for systems ergonomics where a holistic viewpoint is essential, and yet this is the classic approach of the experimental paradigm. This is elaborated next (in a lengthy quotation), in the context of small group research (Arrow et al. 2000), which is characteristic of many systems ergonomics interventions:

... the nomothetic paradigm, as represented in the experimental tradition, makes these assumptions;

1. All cases being treated as alike are indeed alike, in all ways that matter, on all variables except the experimental manipulanda and the dependent variables. Of course, this is frequently not true. The problem is generally handled by the next assumption.

2. Differences or idiosyncrasies that do exist among those cases, on aspects not being studied, will be more or less evenly distributed among experimental conditions, provided cases are assigned to conditions via random allocation, a *sine qua non* for a true experiment.
3. The set of cases studied represents a larger population of cases, with a quantifiable margin of error, provided the cases are selected via random sampling from this defined population (a *sine qua non* for generalisation; resulting values for the cases studied (sample means and variances) can be used to estimate parameters (population means and variances) for the population as a whole.
4. The data collection methods used for any study inevitably contain both systematic and random errors of measurement, but steps such as counter-balancing, use of multiple measures and random allocation of cases to conditions will minimise systematic errors of measurement. Residual errors of measurement are assumed to be random.
5. Random errors of measurement are assumed to be independent of each other and of conditions and are assumed to be normally distributed.

Where people in systems are the subjects of experimentation, not all of these assumptions stand up to close examination. These points are not new; Chapanis provided early illumination of these issues (Chapanis 1967), and though he may not have formally known of the complexity concept, his comments are still valid.

There can be no doubt, however, that there is a role for experimentation, in aspects of physiological investigation or when there is no available real-world system for study, or when the funds and other resources available allow of no other approach.

Alternatives to experimentation do exist, such as naturalistic comparative case studies, simulations and computational models, and these may enable a better understanding of reality.

## STANDARDS IN SYSTEMS ERGONOMICS

A good discussion of standards in general will be found in Chapter 37. For this reason, this chapter is restricted to a mention of some of the important standards within the domain of systems ergonomics that will constrain the selection of methods, given the requirements set out in the standard. Table 33.1 provides a brief description of these. We turn now to a discussion of methods.

## CLASSES OF METHODS IN SYSTEMS ERGONOMICS

Fortunately, there is a multitude of methods for use by systems ergonomists, plucked from a range of disciplines, and covering all aspects of the system life cycle and its support. Unfortunately, there are far too many to outline in this chapter. This has necessitated a different approach; the next section contains a table that lists a number of texts describing individual methods and that could be used as reference texts. This is followed by some exemplar methods for the different phases of the system life cycle, included for illustration purposes only.

## METHODS HANDBOOKS

A surprising number of these exist, and Table 33.2 provides a brief description of some of those available (there are many more). There is no implied ordering of the texts listed.

The actual selection of methods depends both on what the purpose of the exercise is and on the sources of information that are available to you (note that the available sources may not be the right ones, just the ones to which you have access at this point in the space–time continuum).



**TABLE 33.1**  
**Some Important Standards for Systems Ergonomics**

| Title                                                                                                                                                                                                          | Description                                                                                                                                                                                                                                                                                                                                                           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO 6385:2004 Ergonomic principles in the design of work systems                                                                                                                                               | Covers the basics of designing a workstation within the context of systems. Includes aspects of occupational safety.                                                                                                                                                                                                                                                  |
| ISO 900x Quality management                                                                                                                                                                                    | ISO 900x is a family of standards, including<br>ISO 9001:2008 – Sets out the requirements of a quality management system.<br>ISO 9000:2005 – Covers the basic concepts and language.<br>ISO 9004:2009 – Focuses on how to make a quality management system more efficient and effective.<br>ISO 9001:2008 – Sets out the requirements of a quality management system. |
| ISO 9241-210:2010 Ergonomics of human–system interaction                                                                                                                                                       | Part of the ISO 9241 family (see Chapter 37); provides requirements and recommendations for human-centred design principles and activities throughout the life cycle of computer-based interactive systems.                                                                                                                                                           |
| ISO 10006:2003 Quality management systems – guidelines for quality management in projects                                                                                                                      | Provides guidance for projects of varying complexity, small or large, of short or long duration, in different environments and irrespective of the kind of product or process involved. This can necessitate some tailoring of the guidance to suit a particular project.                                                                                             |
| ISO 11354-1:2011 Advanced automation technologies and their applications – requirements for establishing manufacturing enterprise process interoperability – part 1: framework for enterprise interoperability | This is the first part of a multi-part standard intended to ensure interoperability between enterprises for information relating to business, process, services and data. Other parts are still under development.                                                                                                                                                    |
| ISO 12207 Software engineering: software life cycle processes                                                                                                                                                  | Establishes a common framework for software life cycle processes, with well-defined terminology, that can be referenced by the software industry. Shares the same terminology as ISO/IEC 15288, so the two work well together when integrating a software development process.                                                                                        |
| ISO/IEC 15288:2008 Systems engineering: system life cycle processes                                                                                                                                            | Describes all the essential elements that a complete system engineering process should have. Should be read in conjunction with ISO/IEC TR 19760.                                                                                                                                                                                                                     |
| ISO 19011:2011 Guidelines for auditing management systems                                                                                                                                                      | Provides guidance on auditing management systems: principles of auditing, managing an audit programme and conducting management system audits, auditors and audit teams.                                                                                                                                                                                              |
| ISO TR 18529:2000 Ergonomics – ergonomics of human system interaction – human-centred life cycle process descriptions                                                                                          | The usability maturity model in the standard helps to identify the capability of an organisation to apply user-centred design methods, and areas for potential improvement.                                                                                                                                                                                           |
| ISO/IEC TR 19760:2003 A guide for the application of ISO/IEC 15288                                                                                                                                             | A technical report that provides guidance for application of the International Standard ISO/IEC 15288 Systems Engineering – System life cycle processes.                                                                                                                                                                                                              |
| ISO/IEC 19763 Information technology – metamodel framework for interoperability (MFI)                                                                                                                          | Provides the basis for connecting systems together into an SoS. Aimed mainly at the technical issues rather than the organisational issues.                                                                                                                                                                                                                           |

*Note:* A further discussion of standards will be found in Chapter 37.

**TABLE 33.2**  
**Listing of Useful Methods Texts for Systems Ergonomics**

| Text                                                                                                                  | Year | Comment                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------------------------------------------------------------------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Handbook of Human Factors and Ergonomics Methods.</i> Stanton et al., CRC Press, ISBN 0415287006                   | 2005 | Covers over 80 methods; each has a chapter, in a standard layout. Has a macroergonomics section that describes 14 methods directly relevant to systems ergonomics; the rest are fine-grain for aspects of human behaviour.                                                                                                                           |
| <i>Human Factors Methods – A Practical Guide for Engineering and Design.</i> Stanton et al., Ashgate, ISBN 0754646610 | 2005 | Lists over 100 methods, each in a standard layout, including strengths and weaknesses. Has a strong bias towards cognitive assessment of systems. Includes a number of very useful tables to assist in finding appropriate methods. Some 20 methods are directly relevant to systems ergonomics; rest are fine-grain for aspects of human behaviour. |
| <i>A guide to Task Analysis.</i> Amazon Kirwan and Ainsworth, Kindle edition, ASIN: B000Q35U1A                        | 2007 | Re-issue of 1992 book; good source of methods to understand more about real-world systems, and still relevant to current practice.                                                                                                                                                                                                                   |
| <i>Human Reliability Assessment.</i> Kirwan, Taylor & Francis Group, ISBN 0415267579                                  | 2005 | Best source of methods and information on an important aspect of systems ergonomics. However, field is not static, and it would be advisable to look at other texts as well, such as Leveson (2011), Dekker (2002) and Hollnagel et al. (2006).                                                                                                      |
| <i>Research Methods for Business Students.</i> Saunders et al., Pearson, ISBN 0273750755                              | 2012 | Good, all-round text on methods, extending into business domains and concerns; good chapters on subjective methods.                                                                                                                                                                                                                                  |
| <i>Quantitative Methods for Business.</i> Camm et al., S-W College Publishing, ISBN 1133584462                        | 2012 | Provides methodological links to management and engineering disciplines. Indicates usages for systems ergonomics information (or lack thereof).                                                                                                                                                                                                      |
| <i>Handbook of Human Digital Modelling.</i> Duffy et al., CRC Press, ISBN 0805856463                                  | 2009 | As soon as they are published, books on this topic are out of date. Nevertheless, this provides a comprehensive account of modelling, methods, pitfalls, etc.                                                                                                                                                                                        |
| <i>The Occupational Ergonomics Handbook</i> , 2nd ed. Karwowski et al., CRC Press, ISBN 0849326419                    | 2006 | In two volumes; provides a comprehensive coverage of tools and techniques, many of which are suitable for systems ergonomics. As the title indicates, it covers much of the biomechanical and physiological aspects of the topic.                                                                                                                    |
| <i>Human Factors for Engineers.</i> Sandom et al., IEE Press, ISBN 0863413293                                         | 2004 | Presents a wealth of human factors knowledge, mainly in the cognitive domain, succinctly condensed for engineers. Also has several chapters on methods. Good indicator of the interface to engineers.                                                                                                                                                |
| <i>Focus Groups.</i> Langford et al., Taylor & Francis Group, ISBN 0415262089                                         | 2003 | Covers just a single method, but is a very practical discussion of focus groups and their application.                                                                                                                                                                                                                                               |
| <i>Handbook of Human Factors and Ergonomics.</i> 4th ed. Salvendy, Wiley, ISBN 0471449172                             | 2012 | Contains a wealth of knowledge and information, and has brief descriptions of many methods. Other texts would need to be consulted for more detailed information about the methods.                                                                                                                                                                  |
| <i>Handbook of Human Systems Integration.</i> Booher, Wiley, ISBN 0471020532                                          | 2003 | Comprehensive coverage of the systems ergonomics domain, with applications. Part 3, about 300 pages, is about methods, tools and techniques, several of which are not covered in other texts.                                                                                                                                                        |
| <i>Qualitative Data Analysis.</i> Miles et al., Sage, ISBN 0803955405                                                 | 1994 | Old, but still one of the best texts available. Has a bias towards research in the education domain. Detailed descriptions of qualitative methods.                                                                                                                                                                                                   |
| <i>Engineering Systems.</i> de Weck et al., MIT Press, ISBN 0262016702                                                | 2011 | About the engineering of systems; outlines the generic issues well. Chapter 5 discusses relevant tools and methods.                                                                                                                                                                                                                                  |

(Continued)

**TABLE 33.2 (Continued)**  
**Listing of Useful Methods Texts for Systems Ergonomics**

| <b>Text</b>                                                                                                                                                                                                   | <b>Year</b> | <b>Comment</b>                                                                                                                                                               |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Systems Engineering Handbook</i> . Haskins, INCOSE-TP-2003-002-03.2, <a href="http://www.incose.org/ProductsPubs/products/sehandbook.aspx">http://www.incose.org/ProductsPubs/products/sehandbook.aspx</a> | 2010        | Exhaustive outline of systems engineering orthodoxy, outlining processes for all phases of the system life cycle, with chapters on the role of simulation and other methods. |
| <i>Handbook of Systems Engineering and Management</i> . Sage et al., Wiley, ISBN 0470083530                                                                                                                   | 2009        | Edited book, covering all aspects of systems engineering. Has a strong content of systems ergonomics thinking; outlines many useful methods.                                 |
| <i>Verification and Validation of Complex Systems, Human Factors Issues</i> . Wise et al., ISBN 3540565744                                                                                                    | 1993        | As the title says, it addresses the issues of V&V for complex systems. Has a bias towards air traffic control, but methods and techniques are valid for other domains.       |

### SOURCES OF INFORMATION FOR SYSTEMS ERGONOMISTS

Table 33.3 provides a listing of the usual sources accessed by systems ergonomists. The assumption underlying Table 33.3 is that on most occasions the exercise will be undertaken within the context of some organisation.

### METHODS FOR COLLECTING DATA

In this section, we have taken a slightly unorthodox approach to the methods. Normally, one would list the requirements to use the method(s). However, in practice, one finds that one uses any available method that provides reliable, useful data at the time that it is needed, for whatever the purpose is at that time. For this reason, we have discussed the methods from an ‘affordance’ perspective, indicating what the methods can give you, rather than discussing when, where and why one should use them.

Actually collecting the data that you need will obviously require a number of different methods from the sources shown in Table 33.3; however, unorthodox methods may be very useful too. An example comes from Webb, Campbell et al. (1959); in a Canadian museum, they wanted to know which exhibits appealed to children of different age groups, so that they could reposition the displays more usefully for the visitors. In those days, all the exhibits were kept in glass cabinets. So, at the end of the day the investigators went round and counted the number of nose prints on the glass and measured how high each one was. This provided information of sufficient accuracy to enable them to redesign the layout.

This leads to Table 33.4, outlining sources for data of particular kinds of purpose, where these purposes fit into the classes of a systems development process, as shown in Figure 33.5.

There are some universal methods that the systems ergonomics practitioner will always use, from the moment the exercise starts until it ends with the accumulation of lessons learnt (because every study tells the practitioner something new). These basic methods are observation studies, interviews and document searches (whether on paper or through the Internet). These basic techniques are covered in more detail in Chapters 3 to 6 and in the handbooks listed in Table 33.1. They are necessary because they cover the exploration aspects of systems ergonomics, wherein the practitioner finds out what the purpose is and is not, establishes where the system boundary is and uncovers what actually happens in the system as opposed to what the documents say and many people think. Judicious use of these techniques can also shorten greatly the time spent on establishing a good understanding of the system.

**TABLE 33.3**  
**Sources of Information in Systems Ergonomics**

| Type of Information                                               | Source of Information                                                                                                                                                                                                                                                                                                                                                                                       | Comments                                                                                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Strategy and system purposes, which determine the system boundary | Senior managers, formal vision statements, engineering resource plans. Interviews and document searches are fundamental to this.                                                                                                                                                                                                                                                                            | Importance of this information is proportional to the time period for which your final analysis is expected to be relevant.                                                                                                                                                                                               |
| Operational policies                                              | Senior managers, policy documents, codes of conduct, notices to employees. Informal sources include the employees who are affected by these policies. It is important to contact as many stakeholders as possible; there will be many different opinions regarding policies and their execution.                                                                                                            | Policies often conflict, and may be uneven in their application. Frequently, the real-life application of policies involves workarounds and informal deals; while it is necessary to know the formal policies it is always beneficial to check with those responsible for their application how exactly they are applied. |
| Standards                                                         | Internal: design standards, process engineering standards, (including SOPs and SPIs), product engineering standards.<br>External: Supplier proprietary standards, industry standards, national standards, international standards.<br>For both of these, it is the stakeholder's interpretation of these standards in doing the work of the system that is important, not just the wording of the standard. | Note that there may be several conflicting sets, if the process serves customers in different nations, or if different groups of personnel carry out the same process in different locations.                                                                                                                             |
| System constraints                                                | Documents and opinions of those involved in the characteristics of system resources, design of the transformation process, process engineering, health and safety, legal issues, staffing agreements, facilities management, process operations, distribution issues.                                                                                                                                       | Some of these will be documented (e.g. 'do not run this machine faster than ...'); others will be experiential (e.g. 'do those before lunch, if you want them posted today').                                                                                                                                             |
| System SPIs                                                       | Documents, line management, designers, supervisors, operators.                                                                                                                                                                                                                                                                                                                                              | Statements of what needs to be done in order to deliver the system product. These may not have sufficient detail, and likely will not reflect current practice.                                                                                                                                                           |
| System SOPs                                                       | Documents, line management, designers, supervisors, operators.                                                                                                                                                                                                                                                                                                                                              | Statements about how to operate particular machines, applications, tools. May be out of date and may not be correct.                                                                                                                                                                                                      |
| System process characteristics and behaviour                      | Operators, supervisors, managers. There will be many different opinions.                                                                                                                                                                                                                                                                                                                                    | Tends to be held in human heads, not in documents. Beware how old the documents are.                                                                                                                                                                                                                                      |
| System failures                                                   | Maintenance records, Health and safety records, legal records, internal documents, newspapers.                                                                                                                                                                                                                                                                                                              | Records may not be in a useful format for your purposes, nor record the detail you need, nor be accurate.                                                                                                                                                                                                                 |
| System process quality                                            | Process quality records, notes of quality circle meetings, 'Voice of the customer' projects, other stakeholder opinions.                                                                                                                                                                                                                                                                                    | The numbers are important; even more so is the interpretations of these. Ask the question, 'what does the number mean, and for how long?'                                                                                                                                                                                 |

(Continued)

**TABLE 33.3 (Continued)**  
**Sources of Information in Systems Ergonomics**

| Type of Information                                                              | Source of Information                                                         | Comments                                                                                                                                                                                                                                                                                                                 |
|----------------------------------------------------------------------------------|-------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Process limitations                                                              | People, case study references, anecdotes, design documents, operator reports. | Many undocumented and with individual perspectives. Interviews are important to create a cohesive and coherent understanding.                                                                                                                                                                                            |
| Organisational structure                                                         | Line management documents, human resource management documents.               | Formal statements regarding responsibilities and reporting structures, but may not match what actually happens.                                                                                                                                                                                                          |
| Organisational culture and ways of working: 'the way we actually do things here' | People.                                                                       | Almost certainly undocumented, and the biggest determinant of actual process operation – linked to quality, failures and limitations.                                                                                                                                                                                    |
| Benchmarks                                                                       | Can be any combination of the formal documents previously presented.          | A crucial source of information. Other non-competitors in other sectors may have similar processes, and may be best-in-class, with whom you can share knowledge. Competitors may share information as well; for example, in the pharmaceutical industry, competitors share information about the design of pilot plants. |

*Note:* Note the number of different sources and their different natures, and hence the need for using many data capture methods.

However, there are other techniques with which the practitioner ought to be equipped, most of which rely on the data obtained through use of the universal methods given earlier. These tend to fall into the analytic category; dealing with the issues of making sense of the data that is being accumulated (because the picture is seldom simple). We describe some of these later.

## **SYSTEMS DESIGN: METHODS FROM SYSTEMS ENGINEERING AND OTHER DISCIPLINES**

For simple systems, it may be possible just to use the universal methods and rely on one's mental picture of the system and diagrams for further work. However, the increasing interoperation of systems and their increasing complexity is likely to require other methods to understand system structure and behaviour. This takes us into the realms of modelling and simulation, to comprehend both structure and resulting behaviour of the system. A side benefit that comes from this is that the model-building process can reveal where there is suspect data, necessitating a return to data-gathering.

While there are still advantages in physical model-building (e.g. the auto industry), it is likely that more and more systems models will be built in software. This is borne out by the steady growth in importance of model-based systems engineering (MBSE) for both complex systems and for the design of systems of systems. We now discuss some of the methods that are relevant to the understanding of larger systems. However, it should always be borne in mind that any kind of model is an incomplete representation of reality. The implication of this is that by designing a system using models alone, a system will be delivered that has hidden surprises that may reveal themselves later when the system is in operation.

**TABLE 33.4**  
**Classes of Methods for Data Capture**

| Method                       | Comments                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Literature searching         | <p>Nowadays, almost all current, documented information will be found online, given the right permissions. However, given the speed at which storage and retrieval technology progresses, much data may no longer be accessible. For 'old' information (pre-2000), paper-based records are of increasing importance as you look backwards.</p> <p>Records may not be in a suitable structure and may omit important information for your purposes (because of brevity, clipping, condensation).</p> <p>What you find in the professional and technical literature is likely to have been refereed and is likely to be trustworthy.</p> |
| Web-based searches           | <p>Search should involve both the intranet and Internet. Access and a good search engine are vital.</p> <p>What you find may be neither trustworthy nor directly relevant to the context of interest. As has been said of regression analysis, what you find is a basis for further investigations, not an end in itself.</p>                                                                                                                                                                                                                                                                                                          |
| Interviews with stakeholders | <p>The main source of undocumented information. It is always wise to interview as many people as possible, to discover alternative viewpoints, identify errors and to corroborate what you have been told or discovered by other methods. An important aspect is that you get a view from the inside, compared to your view from the outside.</p>                                                                                                                                                                                                                                                                                      |
| Surveys                      | <p>Only useful for benchmarking, for process snapshots and for establishing baseline performance indices. Variations in real-life contexts can vitiate the value of surveys, and low rates of response can render them useless.</p>                                                                                                                                                                                                                                                                                                                                                                                                    |
| Observation                  | <p>Many different classes of observational techniques; all of them need practice.</p> <p>Offer a bottom-up view of the process and can reveal when the real system processes do not match the documents, but do deliver results.</p> <p>What you see is physical; observation provides little guidance on the cognitive processes of people in the system.</p>                                                                                                                                                                                                                                                                         |
| Traceability studies         | <p>Essentially involves following a product, document, datum or decision through the processes of interest within the system.</p> <p>Allows you to check that what is supposed to happen does in fact happen, but may be subject to one-off events during the course of the study.</p>                                                                                                                                                                                                                                                                                                                                                 |
| Simulations or models        | <p>May be the only way to make sense of a complex system. Likely to require frequent iterations of the methods given earlier, until a coherent picture emerges.</p> <p>Techniques can range from virtual reality, 'wizard of Oz' simulations, mock-ups using cardboard boxes and paper-based models as well as the more usual software-based models.</p>                                                                                                                                                                                                                                                                               |

## Requirements Engineering

First, one must understand the requirements that arise from the stated goals and purpose of the system and the identification of its boundaries. There are a number of methods for this, and a brief list is given as follows:

- The stakeholder influence map (Haskins 2010) to provide a display of the relative influence that different individuals and groups have over decision-making and system operation.
- Interviews/questionnaires (Langford and McDonagh 2002, Stanton et al. 2005b). Basic methods for gathering data, and important for discovering what humans in the systems think (customers, users, maintainers or designers).
- Use cases (Holt 2004). Also known as scenario development, a vision of the intended system-in-use is created, from which and by comparison with what exists now, one may develop timelines of steps and the requirements to achieve these steps, typically defining interactions and relationships between roles and a technical system, to achieve a goal.



Roles can be people, programmes or other systems. These act as exemplars of the system at work and are a good means of ensuring a common understanding among people.

- Requirements tree (Robertson and Robertson 2012). A hierarchy of the system's requirements, to enable short-hand discussions and to provide clarity about requirements and their inter-relationships.
- Systemic textual analysis (Robertson and Robertson 2012). A method from the software domain to capture requirements. On the basis of operational requirements (or purposes), one defines functional requirements, and then all the so-called non-functional requirements.
- Viewpoint analysis (Robertson and Robertson 2012). An alternative method for eliciting system requirements, particularly useful when there is limited input from system stakeholders. On the basis of what data have been obtained, an 'expert' group then expands the requirements, making use of their own knowledge.
- Quality function deployment (Akao 1989). A technique first introduced in the automotive industry. In its full form, it commences with stakeholder interviews, and then enables these to be transformed successively, with the help of additional expert knowledge, into system requirements, then system components to meet those requirements, and then the supply chain to deliver the system.
- 'Voice of the customer' (Yang 2007). A collective title for obtaining stakeholder opinions about current systems or analogous systems to that being planned. Various methods can be used, such as walkthroughs, focus groups, interviews and so on.

### Network Models

For systems ergonomics, the most useful network models are those concerned with social networks, in which the most common approach seems to be social network analysis (SNA) (Scott 2012). This is very much concerned with the relationships (links) between individuals (or groups, or organisations, treated as single entities). The nodes of the network are often treated simplistically because the focus is on the interactions. The premise is that it is the interactions which determine both the future structure of the network and its overall behaviour. A number of useful measures have arisen from this work, such as maximum and minimum spanning trees for the transfer of information and the notion of network centrality (or importance).

A note of caution is advisable here; the connections in the networks say nothing about the meaning of what was communicated, nor its timeliness, nor its correctness.

### Discrete Event Simulation

This class of techniques is predicated on the notion of a queue of events which occur typically in series, and then measuring the system's response to the occurrences (Duffy and Salvendy 2008). These events may be external inputs to the system or alterations to system components themselves (e.g. a module failure). The key components in this class are the universal time-clock, the queue of events, the random number generator for times to release events into the system, the simulation model of the system and the statistical tracking of system behaviour. The approach is useful for exploring rarely encountered situations that might cause system failures (particularly where frequency and timeliness are important) or are unlikely to occur in observational studies.

### Agent-Based Modelling

This class of models is for the simulation of the actions and interactions of agents (or groups or organisations) operating in parallel in a shared scenario, and measuring both the overall behaviour and the behaviour of individual agents within the scenario (Duffy and Salvendy 2008). The agents themselves may be elaborate and have complex behaviour such as reproduction, reorganisation and so on. This is an approach suited to larger, more complex exercises, such as in SoS studies; however,

it can become very expensive in terms of manpower, time, skills and money in setting up and running the model, and also in interpreting what the model displays.

On the other hand, it is about the only way that complex systems essential to society can be explored before they are instantiated. In particular, it allows the creation of 'black-box' models, in which independent organisations in an SoS model the important aspects of their participating system while retaining confidentiality about it, and link them together to create behavioural models of the whole SoS.

### **System Dynamics**

This class of models is concerned mainly with the effects of feedback in systems and the resulting behaviour of the system (Duffy and Salvendy 2008). The main ingredients of these models are feedback loops, time delays, stocks and flows. Typically, a system is modelled and initial conditions are set, with stocks at various places in the system. The model is then allowed run in small incremental steps, and the flow of stocks through the system is recorded. The important characteristic and reason for these models is that seemingly simple static models of systems can reveal unexpected dynamic behaviour.

For all of these model classes, an important aspect is dependency analysis; each constituent agent in a model will have its own idiosyncrasies and variations from the real equivalent, all of which lead to inaccurate predictions and performance measures. The problem comes when these small-scale errors are aggregated and compound themselves in large-scale models to produce major errors in model performance.

### **Human Cognitive Performance Modelling**

A number of the texts listed in Table 33.2 provide as many methods as one might ever require; however, a particularly good source of methods for this topic is Stanton et al. (2005b). However, the vast majority of the methods listed are appropriate for the individual; there is a lack of methods that consider teamwork, reflecting a wider problem for systems ergonomics. There is a chapter devoted to team assessment, but this deals mainly with task analysis and communications, based on observational data captured by pen and paper methods.

Other chapters in this book, as in most of the texts listed, cover topics such as task analysis and cognitive task analysis, human error identification, situation awareness, mental workload assessment and interface analysis.

### **Process Analysis**

A key feature of most systems is the process by which the system achieves its purpose(s). In the vast majority of systems, there are defined processes, expected to be repeatable. Consequently, process analysis methods are of great importance to systems ergonomics practitioners.

There is a large variety of methods; some developed with systems ergonomics in mind (e.g. hierarchical task analysis [HTA]), and others from other cognate disciplines (e.g. Petri nets, State Charts and IDEF, from the software domain). In all of these, the intention is to capture what happens in the process; the output varies, depending on the method and the purpose of the study. For example, IDEF captures the sequence and parallelism of tasks in time, whereas Petri nets and State charts enable the analyst more easily to see the state of the system at a given time. HTA, discussed in more detail in Chapter 6, gives good hierarchical decomposition and vision of tasks down to the finest grain deemed necessary. All of the methods capture the four essential ingredients of any process task: inputs, outputs, resources required to carry out the transformation (e.g. tools, containers, manuals) and controls on the task (starts, stops, pauses).

Since most of these methods rely on much the same set of process primitives, it would be handy if there was a single tool to capture the raw data and then develop the process analysis output that the analyst requires as a post-processing set of alternatives. Unfortunately, this does not exist, perhaps for good reasons.

### Physiological Assessment Methods

A few of the texts listed in Table 33.2 outline methods with a physiological flavour; Chapters 15 (physiological assessment) and 16 (physical assessment) in this book give some further discussion. These are important for systems ergonomics because the personnel involved in systems may, over a period of time, suffer physiological effects from their work. A good text (among several) in which to find many methods discussed is Stanton et al. (2005a) covering psychophysiological methods; musculoskeletal, physiological and physical performance methods, and methods for the measurement of the human physical environment in which work takes place. As a generalisation, these methods are concerned with habitability for humans executing the system processes, performance decrements and the effects of cognitive and physiological fatigue as induced by environmental factors. In the light of moves to 12 hour shifts and 7-day working, these methods are important contributors to systems ergonomics.

While some methods in these classes are observational for E/HF practitioners in the field, most require sensors and tooling to capture the relevant data. A problem with this is that this may interfere with the normal performance of the operators who are being assessed, even if the sensors and tooling are not directly attached. A further problem is that sensors sense only what they sense; inferring cognitive states from psychophysiological data may not be accurate, for example.

### Other Useful Techniques

It is worth mentioning a number of methods that may be useful for particular purposes. One such is verbal protocol analysis for the elicitation of expertise (Chapter 7 of this book), important for the development of roles and control in systems, and role modelling for team building in systems (Chapter 29 of this book). The understanding of entities can also be assessed, for example, by means of repertory grid techniques, which can provide insights into the emotional aspects of systems as experienced by the human stakeholders involved in them.

## CONCLUSIONS

It will have become evident that there are enough methods available to systems ergonomics practitioners to cover almost all situations that might be encountered; the problem is to choose those that would fit the exercise being undertaken, both with respect to the purpose of the exercise, the predilections of the client and the resources that are available – the latter being time, money, manpower and equipment, in the main. There are no rules for this; the choice is always contingent on the particular circumstances.

Secondly, we have chosen to list classes of methods in this chapter, again because of the multitude of methods from which to choose, and then provide sources for more information. Within the confines of this chapter, this is perhaps the most useful approach that we could adopt.

Finally, we hope that you find this chapter useful. It is perhaps best read by dipping in and out of the other chapters cross-referenced in the text.

## REFERENCES

- Akao, Y. (1989). *Quality Function Deployment: Integrating Quality Requirements into Product Design*. Cambridge, MA: Productivity Press.
- Allen, P. M. (1998). Evolving complexity in social science. In *Systems: New Paradigms for the Human Sciences*, G. Altman and W. A. Koch (eds.). New York: Walter de Gruyter.
- Allen, P. M., J. Boulton, M. Strathern and J. Baldwin (2005). The implications of complexity for business process and strategy. In *Managing Organisational Complexity: Philosophy, Theory and Application*, K. Richardson (ed.), pp. 397–418. Mansfield, MA: ISCE Publishing.
- Allwood, J. M., M. F. Ashby, T. G. Gutowski and E. Worrell (2011). Material efficiency: A white paper. *Resources, Conservation and Recycling* **55**: 362–381.

- Andersson, G., P. Donalek, R. Farmer, N. Hatziaargyriou, I. Kamwa, P. Kundur, N. Martins et al. (2005). Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. *IEEE Transactions on Power Systems* **20**(4): 1922–1928.
- Arrow, H., J. E. McGrath and J. L. Berdahl (2000). *Small Groups as Complex Systems: Formation, Co-ordination, Development, and Adaptation*. London, U.K.: Sage Publications.
- Asaro, P. M. (2011). Remote-control crimes. *IEEE Robotics & Automation* **18**(1): 68–71.
- Asaro, P. M. (2012). A body to kick, but no soul to damn: Perspectives on robotics. In *Robot Ethics*, P. Lin, K. Abney and G. A. Bekey (eds.), pp. 169–186. Cambridge, MA: MIT Press.
- Ashby, W. R. (1956). *An Introduction to Cybernetics*. London, U.K.: Chapman & Hall.
- Azani, C. H. (2008). System of systems architecting via natural development principles. In *IEEE International Conference on System of Systems Engineering, 2008 (SoSE '08)*, pp. 1–6. Singapore: IEEE.
- Barot, V., S. Henson, C. E. Siemieniuch, M. J. D. C. Henshaw, M. A. Sinclair, S. L. Lim, C. Ncube, M. Jamshidi and D. DeLaurentis (2012). State of the art report: Trans-Atlantic Research and Education Agenda in Systems of Systems (T-AREA-SoS). In *p. n. EU Framework Programme 7*, p. 126. Loughborough, U.K.: Loughborough University.
- Bertalanffy, L. V. (1968). *General Systems Theory: Foundations, Development, Applications*. New York: George Braziller.
- Bruseberg, A. and G. Fletcher (2008). *The Human View – Handbook for MODAF*. Bristol, U.K.: Systems Engineering & Assessment Ltd.
- Campbell, D. T. and A. Fiske (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychology Bulletin* **56**(2): 81–105.
- Chapanis, A. (1967). The relevance of laboratory studies to practical situations. *Ergonomics* **10**(5): 557–577.
- Checkland, P. B. (1995). Soft systems methodology and its relevance to the development of information systems. *Information Systems Provision: The Contribution of Soft Systems Methodology*, F. A. Stowell (ed.), pp. 1–17. Maidenhead, U.K.: McGraw-Hill.
- Cilliers, P. (2005). Complexity, deconstruction and relativism. *Theory, Culture & Society* **22**(5): 255–267.
- Conant, R. (ed.) (1981). *Mechanisms of Intelligence: Ross Ashby's Writings on Cybernetics*. Seaside, CA: Intersystems Publishers.
- Dahmann, J. and K. Baldwin (2008). Understanding the current State of US defense systems of systems and the implications for systems engineering. In *2nd Annual IEEE Systems Conference*, Montreal, Quebec, Canada.
- Davenport, G. (1979). *Herakleitos and Diogenes*. Bolinas, CA: Grey Fox Press.
- Dekker, S. W. A. (2002). *The Field Guide to Human Error Investigations*. Aldershot, U.K.: Ashgate Publishing Ltd.
- DTI. (1992). CALS; the executive guide. UK CALS Industry Council, Department of Trade and Industry, U.K.
- Duffy, V. G. and G. Salvendy (2008). *Handbook of Digital Human Modeling*. Boca Raton, FL: CRC Press.
- Edmonds, B. (2000). Complexity and scientific modelling. *Foundations of Science* **5**: 379–390.
- Haskins, C. (2010). *Systems Engineering Handbook: A Guide for System Lifecycle Processes and Activities*. San Diego, CA: International Council on Systems Engineering.
- Heracleitus (535–475 BCE). (cited in G. Davenport, 1979). *Herakleitos and Diogenes*. Bolinas, CA: Grey Fox Press.
- Hitchins, D. K. (2003). *Advanced Systems Thinking, Engineering and Management*. Norwood, MA: Artech House, Inc.
- Hollnagel, E. and D. D. Woods (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton, FL: Taylor & Francis Group.
- Hollnagel, E., D. D. Woods and N. Leveson (eds.) (2006). *Resilience Engineering*. Aldershot, U.K.: Ashgate Publishing Ltd.
- Holt, J. (2004). *UML for Systems Engineering*. London, U.K.: IEE Publishing.
- Jamshidi, M. (ed.) (2009a). *System of Systems Engineering: Innovations for the 21st Century*. Hoboken, USA: John Wiley & Sons.
- Jamshidi, M. (ed.) (2009b). *Systems of Systems Engineering: Principles and Applications*. Boca Raton, FL: CRC Press.
- Kirwan, B. (2005). Human reliability assessment. In *Evaluation of Human Work*, J. R. Wilson and E. N. Corlett (eds.), pp. 833–877. London, U.K.: Taylor & Francis Group.
- Langford, J. and D. McDonagh (2002). What can focus groups offer us? *Contemporary Ergonomics 2002 – Proceedings of the Annual Conference of the Ergonomics Society*, P. T. McCabe (ed.). London, U.K.: Taylor & Francis Group.

- Leveson, N. G. (2011). Complexity and safety. *Complex Systems Design and Management*, O. Hammami, D. Krob and J.-L. Voirin (eds.), pp. 27–39. Heidelberg, Germany: Springer Verlag.
- Maier, M. W. (1998). Architecting principles for systems-of-systems. *Systems Engineering* **1**(4): 267–284.
- McMichael, A. J. and K. B. G. Dear (2010). Climate change: Heat, health, and longer horizons. *Proceedings of the National Academy of Science* **107**(21): 9483–9484.
- NCOIC. (2011). *NCOIC Interoperability Framework 2.1*. Washington, DC: Network Centric Operations Industry Consortium.
- Ohren, O. D. (2004). An ontological approach to characterising enterprise architecture frameworks. In *ICEIMT '04*, Toronto, Ontario, Canada.
- Perrow, C. (1999). *Normal Accidents: Living with High-Risk Technologies*. Princeton, NJ: Princeton University Press.
- Rasmussen, J. (1994). Complex systems, human factors, and design of teaching curricula. In *Mensch Machine Systeme und Neue Informationstechnologien*, K. P. Gärtner, W. Stein and H. Widdel (eds.). Aachen, Germany: Verlag der Augustinius Buchhandlung.
- Rasmussen, J. and L. P. Goodstein (1986). Decision support in supervisory control. In *Analysis, Design and Evaluation of Man-Machine Systems, 2nd IFAC/IFIP/IFORS/IEA Conference*, Varese, Italy.
- Reason, J. T. (2008). *The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries*. Farnham, U.K.: Ashgate Publishing Ltd.
- Richardson, K. A. (ed.) (2005). *Managing Organisational Complexity*. Greenwich, CT: Information Age Publishing.
- Robertson, S. and J. Robertson (2012). *Mastering the Requirements Process: Getting Requirements Right*.
- Romero, J. (2012). Lack of rain a leading cause of Indian grid collapse. *IEEE Spectrum*. New York: IEEE.
- Scott, J. P. (2012). *Social Network Analysis*. San Francisco, CA: SAGE.
- Senge, P. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organisation*. New York: Doubleday.
- Sherwood, S. C. and M. Huber (2010). An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences* **107**: 9552–9555.
- Siemieniuch, C. E. and M. A. Sinclair (2002). On complexity, process ownership and organisational learning in manufacturing organisations, from an ergonomics perspective. *Applied Ergonomics* **33**(5): 449–462.
- Sinclair, C. E. and M. A. Siemieniuch (2005). Ergonomics aspects of ‘immortal’ systems. In *Contemporary Ergonomics 2005*, P. D. Bust and P. T. McCabe (eds.), pp. 185–189. Leiden, the Netherlands: Taylor & Francis Group.
- Skyttner, L. (1996). *General Systems Theory: An Introduction*. Basingstoke, U.K.: McMillan.
- Sowa, J. F. and J. A. Zachman (1992). Extending and formalising the framework for information systems architecture. *IBM Systems Journal* **31**(3): 590–616.
- Stanton, N., A. Hedge, K. Brookhuis, E. Salas and H. Hendrick (2005a). *Handbook of Human Factors and Ergonomics Methods*. Boca Raton, FL: CRC Press.
- Stanton, N., P. Salmon, G. Walker, C. Baber and D. Jenkins (2005b). *Human Factors Methods: A Practical Guide for Engineering and Design*. Farnham, U.K.: Ashgate Publishing Ltd.
- Sulston, J. (2012). *People and the Planet*. London, U.K.: The Royal Society.
- Sun Microsystems. (1991). CALS and you. Mountain View, CA: Sun Microsystems Federal, Inc.
- Thomas, R. J. (1994). *What Machines Can't Do: Politics and Technology in the Industrial Enterprise*. Berkeley, CA: University of California Press.
- Weick, K. E. and K. M. Sutcliffe (2007). *Managing the Unexpected: Resilient Performance in An Age of Uncertainty*. San Francisco, CA: Jossey-Bass.
- Weick, K. E., K. M. Sutcliffe and D. Obstfeld (1999). Organizing for high reliability: Processes of collective mindfulness. *Research in Organisational Behaviour* **21**: 23–81.
- Woods, D. D. and E. Hollnagel (2006). *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. Basingstoke, U.K.: Taylor & Francis Group.
- Yang, K. (2007). *Voice of the Customer: Capture and Analysis*. New York: McGraw-Hill.
- Yang, Y., B. Boehm and D. Port (2005). A contextualized study of COTS-based e-service projects. In *Fourth International Conference on COTS-Based Software Systems – ICCBSS 2005*. Bilbao, Spain: Springer.





# *Section VI*

---

## *Implementation and Analysis of Human Factors*

The penultimate section of this book presents tools and knowledge that can help us take a top-level perspective to the implementation and analysis of ergonomics/human factors (E/HF). These chapters, three of which are new to this book, should provide the E/HF practitioner with additional knowledge to support both effective delivery of an E/HF programme and a strong justification for the inclusion of E/HF in any good management of a large organisation.

Chapter 34, by Gyi, Shalloe and Wilson, provides a detailed overview of how we can achieve a successful participatory ergonomics approach. This builds on tools that have been mentioned in other chapters of this volume, particularly Chapters 10 and 11, to consider not only how to involve users in design but to ensure that the requirements of all stakeholders are captured in a manner that ensures that ergonomics knowledge can be fed into design requirements and specifications as effectively as possible. One of the things that is also important about the participatory approach is that it is often fun and engaging – working as a group to critique a design, or come up with a new concept is often an enjoyable experience, and many of the tools and approaches that are described in this chapter are designed to ensure that as wide a range of people as possible are able to actively contribute. This ‘fun’ element to the approach should not be ignored, it has an intangible but valuable effect of making participants feel appreciated and respected within their work context, and, as long as hopes are not raised too high, tend to lead to higher acceptance of future changes in the workplace.

Chapter 35 tackles the tricky challenge of cost benefit analysis of E/HF. Tompa and Foley use the case of musculoskeletal disorders to demonstrate how we can measure and elicit the effect of E/HF interventions on employee health and well-being. Cost benefit faces three particular difficulties in the context of E/HF; firstly, many of the variables in which we are interested (such as improved culture or changed work practices) can be difficult to quantify; although quantification is not always appropriate or required for the E/HF practitioner to effectively analyse a situation, quantified data can be easier to use if we wish to make an ‘argument’. I often compare myself to my engineering colleagues in this respect – if they wish to present an argument for investment in a new material for an aircraft fuselage, they can quantify the increased materials and production cost, and demonstrate the offset in aircraft fuel and maintenance cost (acknowledging the presence of other design and environmental variables of course). The situations that ergonomists encounter are rarely this

(relatively) neat and tidy! A second difficulty for cost benefit of E/HF is the low frequency of measurable events, such as incidents. This has led to a focus on capture of near-miss occurrences as well as reportable incidents, but establishment of an effective culture to support near-miss reporting is not always straightforward (as discussed by Ryan in Chapter 31 of this book). The final difficulty is the long-term consequence of E/HF intervention; data such as sickness absenteeism and staff retention can be very powerful, but take time to collect. This tends to be achievable if E/HF is embedded within an organisation, but can be harder to achieve if the E/HF is delivered via an external partnership who may not be able to conduct longitudinal analysis of the effectiveness of any interventions or changes.

Eklund and Yeow (Chapter 36) describe the relationship between ergonomics and quality. Quality is an approach that has been embraced by management science for many years, and it is helpful to understand how the terminology and approaches of E/HF are compatible with the field of quality. The authors of Chapter 36 argue for an integration of the two approaches.

Finally for Section VI, Gray and Stewart (Chapter 37) present an overview of the relevant standards in E/HF design and evaluation. This is a much-needed addition to this text that we hope will provide an invaluable reference to those who study and practice E/HF. An appreciation of the standards to which organisations and industries must adhere is essential for those who are practising E/HF, either as a consultant, or embedded within an organisation, and the authors provide an excellent overview of the wide set of standards which are useful for E/HF practitioners to be aware of. They also encourage readers to get involved with the production of standards, on a national and international level. This is an important request; those of use in all contexts of E/HF, whether conducting fundamental research or implementing new designs, have a contribution to make to the content of future standards, and it is our responsibility as a community to make sure that E/HF considerations are reflected wherever possible and appropriate.

**Sarah Sharples**

---

# 34 Participatory Ergonomics

*Diane Gyi, Sally Shalloe and John R. Wilson*

## CONTENTS

|                                                    |     |
|----------------------------------------------------|-----|
| Introduction.....                                  | 884 |
| Participation and Ergonomics .....                 | 885 |
| Arguments for and against Participation.....       | 887 |
| Methods Used in Participatory Ergonomics .....     | 889 |
| Design Decision Groups .....                       | 892 |
| Extension of the DDG Process.....                  | 892 |
| Participatory Ergonomics Framework .....           | 893 |
| Requirements for Participatory Initiatives .....   | 895 |
| Conclusions.....                                   | 896 |
| Appendix: Participatory Ergonomics Case Study..... | 897 |
| Design Decision Groups Procedure.....              | 897 |
| Scene Setting.....                                 | 897 |
| Session 1 .....                                    | 898 |
| Introduction .....                                 | 898 |
| Word Map.....                                      | 898 |
| Round-Robin Questionnaire.....                     | 898 |
| Drawing of Their Own Workplace .....               | 899 |
| Slide Presentation of Ergonomics Problems .....    | 899 |
| Equipment Demonstration.....                       | 899 |
| Individual Silent Drawing .....                    | 899 |
| Individual Presentation and Discussion .....       | 900 |
| Session 2 .....                                    | 900 |
| Drawing in Pairs.....                              | 900 |
| Presentation and Discussion.....                   | 900 |
| Simple Layout Modelling.....                       | 900 |
| Discussion .....                                   | 900 |
| Fitting Trials .....                               | 901 |
| Session 3 .....                                    | 901 |
| Building a Mock-Up .....                           | 901 |
| Discussion .....                                   | 901 |
| Modifications.....                                 | 901 |
| Simulation Trial and Assessment .....              | 901 |
| Debriefing.....                                    | 901 |
| Outcomes .....                                     | 901 |
| References.....                                    | 902 |

## INTRODUCTION

It is not surprising that throughout this book the issue of participation is raised, both implicitly and explicitly. In a sense it is impossible to carry out 'evaluation of human work' without at least some participation from people actually doing the work. However, when we are talking about participation within systems and product design (Chapters 9 and 11) investigations and interventions studies (Chapter 2) or systems implementation (Chapters 28, 32 and 33), we need something more than user surveys or trials or the direct observation of people at work. This chapter concerns the area that has come to be known as participatory ergonomics (PE). Since the 1980s (although some suggest earlier), there has been a growing interest in the means by which 'non-experts' can become involved in applying ergonomics in different applications. This interest is evidenced in part by the increasing numbers of journal articles, reports and conference papers on the topic, indicating that it has become widespread in both research and practice. This chapter updates the versions of this chapter in previous editions of this text, keeping the historical perspective but also drawing on the current literature and experience of the authors.

If we look in more detail what is meant by PE, several authors are often referenced in the literature. Still widely quoted is Wilson's (1995a) statement that PE concerns 'the involvement of people in planning and controlling a significant amount of their own work activities, with sufficient knowledge and power to influence both processes and outcomes in order to achieve desirable goals' (p. 1071). An early definition is also provided by Nagamachi (1995) for whom PE is 'the worker's active involvement in complementary ergonomics knowledge and procedures in their workplace... supported by their supervisors and managers in order to improve their working conditions and product quality' (p. 371). Eason (1995) broadens the topic to 'establishing design processes in which the end users themselves can influence the design so that it is compatible with their goals and beliefs' (p. 1667). Finally, Kuorinka (1997) has a simpler definition and describes PE as 'practical ergonomics with the participation of necessary actors in problem solving' (p. 268). This is an important definition as it allows for the involvement of all stakeholders in the process. In understanding these definitions, it is important to emphasise that it is not suggested that participants always have a 'blank cheque'; there will almost always be limits on what they can do and on the suggestions they would like to implement. However, in order to do justice to the participatory process and to manage the expectations of participants, certain points will need to be clarified. Limits on what is achievable within the resource constraints set for both the process and the outcome, and thus deciding the main goals and what is needed for participants to be sufficiently influential should ideally be made clear at the outset.

In examining the PE literature, it could be argued that the growth of PE has largely been achieved through 'doing'. A high proportion of the literature concerns reports of PE endeavours in which different researchers describe the execution of one participatory project or another. Whilst this growth may well have served to establish the term within the ergonomics community, it has also led to the term being used to cover a fairly broad range of different approaches and ideas. However, it is not the intention in this chapter to explore the philosophical or theoretical basis for participation at work. Nor will the vast amount of work on employee participation in the arena of industrial democracy, cooperatives or even in a wider political sense be touched on other than indirectly. The concentration is upon participative programmes and techniques which are of immediate relevance to mainstream ergonomics/human factors (E/HF) – that is in the analysis, design, implementation and evaluation required to improve an individual's tasks, equipment, interfaces, workplaces, jobs and work organisation. For a discussion of some of the wider issues in participation, the reader is referred to Mumford (1991), Cotton (1993), Glew et al. (1995) and Riordan et al. (2005).

In this chapter, the development and growth of PE as a discipline is considered along with the potential advantages and disadvantages associated with taking a participatory approach. The focus then moves on to the methods used within PE, before a framework is presented along with a discussion of the requirements for a successful participatory ergonomics intervention. Finally, in an

appendix to this chapter, a case study undertaken by one of the authors is included in order to illustrate the application and methods of PE in more detail.

## **PARTICIPATION AND ERGONOMICS**

There can be little doubt that PE has enjoyed a steady growth due to influences such as human-centred work systems with emphasis upon an involved, responsible workforce; the recognition that motivation and performance are complex issues; increased trade union acceptance of participation; more decentralised and 'flatter' IT-based organisations; greater emphasis on the attributes of quality products, flexibility and good customer service; and the continued introduction of new technical systems and methods of organising them. Indeed, it can be argued that a much more participative culture has developed within many countries. Many of today's high-performing companies are reported to have incorporated employee empowerment and teamwork into their working practices as well as 'people involvement' in processes. For example, both Google and United Airlines foster a culture of employee creativity and innovation to improve company effectiveness by actively encouraging employees to put forward innovative ideas for management to then develop into successful ventures. Also, Unipart runs an in-house programme enabling employees to work in teams to solve problems at their own level.

There have been several excellent systematic reviews of participatory ergonomics in different industries and contexts in recent years (e.g. Hignett et al., 2005; Rivalis et al., 2008; Van Eerd et al., 2010). Hignett et al. (2005) summarise PE interventions in a wide range of industries including health care, manufacturing, construction and transport. They capture interventions at both a micro level, where workers use their knowledge and skills to address ergonomics problems in the workplace, and at a macro level, concerning the systems in place at an organisational level for both strategic direction and operational processes. Although most PE interventions had a micro component, the majority were underpinned by a macro framework. It is also of interest that they found cultural differences in participatory approaches: in the United States, approaches were more at a macro level, whereas in Europe the approach was more diverse. Rivalis et al. (2008) looked at the effectiveness of PE on improving worker health; they found that the majority of the studies they examined reported a positive effect on health outcomes such as absenteeism, musculoskeletal symptoms, lost time injury, sick leave and compensation claims in a wide range of workplaces. However, they indicated that it was important to improve rigour in the study design (e.g. good documentation, use of control groups, random allocation to intervention, etc.) in order to strengthen the evidence base for policy and practice. Finally, Van Eerd et al. (2010) looked at the facilitators and barriers to the implementation of PE interventions in the workplace involving several countries (e.g. the United States, Canada, Europe, Japan) and industry sectors, including manufacturing, agriculture, construction, public administration and health care. They captured elements of successful PE (described later in this chapter) and conclude that addressing key facilitators and barriers such as communication and resources is paramount.

Many PE projects involve setting up a 'team', who are then trained by an expert and, with their new knowledge, then instigate changes in their organisation. Working together in this way to improve ergonomics in the workplace can have a positive impact on worker health and well-being. Van Eerd et al. (2010) comment on the importance of the right balance of expertise in the PE team, for example workers, supervisors, specialists or advisors (from within or outside the organisation). Vink et al. (2006) advocate that for PE activities to be successful, the participation of relevant stakeholders, such as ergonomists, designers, engineers and occupational health professionals, is important and exactly which participants are involved depends on the context and the goal of the intervention.

Developments in European ergonomics and health and safety legislation are also germane to PE. The European Framework Directive (2007/30/EEC) contains specific provisions about consultation and participation of workers; for example it must be ensured that 'the planning and introduction of new technologies are the subject of consultation with the workers and/or their representatives...'

Legislation requires employers to consult employees and especially trade unions, where they are recognised, about matters that affect their health and safety. Furthermore, the need for user/worker involvement and consultation is also emphasised in guidance and standards. For example, in the United Kingdom, publications include 'Managing for Health and Safety' (HSE, 2013); 'Upper limb disorders in the workplace' (HSE, 2002); and British Standard BS OHSAS18001:2007 'Occupational Health and Safety Management Systems – Requirements' (BSI, 2007).

It is increasingly recognised that there is a need for organisations to be more self-reliant in ergonomics. However, many organisations simply do not have the resources to have a consultant ergonomist examine every one of their work processes (even if there were enough ergonomists to go around). This approach does not aid the concept of a health and safety culture of which internal ownership of E/HF can form an important part.

Looking beyond these wider influences, for ergonomists, participation as a philosophy and a process has always been of interest due, in part, to the strong E/HF emphasis on application. People-centred approaches are endemic, from user trials in consumer product testing to user-centred design in human-computer interaction (see Chapter 13) to human-centred technology and sociotechnical programmes in industry (Chapters 28 and 32). Most ergonomists also take a broad view of work redesign and of improving the work environment, arguing that the physical aspects of work neither can nor should be divorced from a consideration of psychosocial and organisational factors. Even when a work redesign programme establishes physical factors as being of greatest concern, to ignore other aspects of work may lead to only partial acceptance and success of any changes made due to the central role of workforce attitudes (Wilson and Grey, 1990).

Participatory initiatives can operate at a number of different levels. In the literature, we can find examples ranging from one-off interventions tackling workplace or equipment design (e.g. see Dos Santos et al., 2011; Parimalam et al., 2012) through to those involving multiple interventions, such as those associated with quality circles or continuous improvement schemes (e.g. see De Looze et al., 2000; Määttä, 2007; Lallemand, 2012). Also, there are full participatory ergonomics management programmes (e.g. see Loisel et al., 2001; Driessen et al., 2008; Guimarães et al., 2014) and even participatory initiatives concerned with organisational systems influencing an entire industry sector (Kardborn, 1998; Moore and Garg, 1998; De Jong and Vink, 2002).

In order that workforce participation in the application of ergonomics methods is most successful, there must be a spreading of relevant skills and confidence throughout the company. There are growing numbers of efforts to train the workforce in E/HF. This training might be aimed at any or all of shop-floor workers, supervisors, industrial, production or design engineers, health and safety personnel, but the most complete programmes certainly include shop floor staff in order to help them assess work and workplaces and propose solutions for E/HF problems. As with E/HF generally, the limited resources available to do this must be recognised, and as a consequence we might establish 'cascaded' training and awareness programmes, which are to some extent further extensions of a 'train the trainers' approach (e.g. see Silverstein et al., 1991; Gyi et al., 2012). Gyi et al. (2012) successfully used a 'train the trainer' approach to raise awareness of the risks to musculoskeletal health of the day-to-day activities of business drivers. Interventions were developed within four case study organisations regarding using the car as an office, manual handling from the car and making drivers more aware of the E/HF issues when choosing a car for work. In essence, to train the trainers meets the need to spread ergonomics more widely, including to non-specialists. Typically, ergonomists train a core team in ergonomics and its education techniques, the training being largely practically based and emphasising 'learning by doing'. Initial training is then passed on to others in the workforce by this core team.

An important very early series of studies by a group of French ergonomists followed the line of providing tools, training, knowledge and confidence for workers to analyse their own work problems and to develop feasible alternatives. The focus of these studies varied from chemical plant control room design including display specifications (Boel et al., 1985), to posture and health analyses



(Montreuil and Laville, 1986), to introduction of new shift work systems (Teiger and Laville, 1987). Their approach was to set up study groups comprising workers, ergonomists and technical staff; the worker representatives were equipped with the tools and skills needed to analyse and make recommendations about particular problems. In some of their work, a task force was formed to oversee the whole project, keeping all levels in the company in touch with current developments. The researchers succeeded in achieving recognition amongst the workforce of the value of an ergonomics perspective and problem-solving process; through their involvement in the administration of questionnaires and direct observation, workforce groups managed to engage the interest of many colleagues in the issues and in finding solutions.

More recently, in 2007, BMW realised that the average age of their production workers was increasing and used a participatory approach for improving health at work and productivity (Loch et al., 2010). As part of a company-wide health awareness initiative, they set up a project with a test production line of workers with an average age of 47 (the 2017 production line) to show that productivity declines are not inevitable with ageing. Initially, there was resistance to the project, but they encouraged these workers to take part in workshops with their line managers (supported by experts such as an ergonomist and a safety officer) to discuss improving well-being at work. Overall, workers came up with 70 load-easing changes many of which were low/no cost – these were small changes to improve the ergonomics of their equipment and workstation which would reduce physical strain on the body. Examples include large handled tools, flooring to dampen pounding, large typefaces and orthopaedic footwear. Management commitment and prompt implementation of ideas encouraged the engagement of workers and snowballing of ideas. The 2017 line was a success for BMW and within 1 year the line was as productive as lines staffed by younger workers, quality targets were achieved and absenteeism had dropped below average for the plant. Many researchers have been inspired by this research, for example, as part of a new dynamics of ageing project (Haslam et al., 2013), an online resource was created to facilitate ways of encouraging managers and workers to think about their body at work and healthy ageing through good design and ergonomics at work.

## ARGUMENTS FOR AND AGAINST PARTICIPATION

As the field of PE has developed, the potential gains associated with using this approach have been reported and discussed. These gains may be direct or indirect (to some extent systemic). It has long been suggested that end user participation in the design of equipment or workplaces will lead to the development of better design solutions, as the practical experience and expertise of end users is brought into the process (Imada, 1991; St-Vincent et al., 1997; De Looze et al., 2000; Vink et al., 2006). However, it is important to clarify what is meant by ‘better’ design as, depending upon the product, the context and the participants, this may mean quite different things. For example, in the case of a piece of industrial machinery, better design can mean one that is more efficient, produces items of a higher quality, is safer to use, is more easy to use or may be used by a wider range of the working population; better design for a consumer product may be entirely to do with usability or aesthetics for example.

As well as better design solutions, the second main advantage for participative processes is said to be greater acceptability of these solutions in the stakeholders (De Jong and Vink, 2002; Gyi et al., 2012). The reasoning is that if people (or their peers) have been involved in generating a solution for change then they are likely to be more committed to making the change work, to be less resistant to change and to be more satisfied as a result (e.g. Wilson, 1995b). If so, and given also that we expect a better designed solution, more fit for purpose, then the process of implementation of change should be more effective and of higher quality. Involvement in a development or implementation process can mean faster and deeper learning of a new system and hence decreased training costs and improved performance (e.g. see Wilson and Grey Taylor, 1995; Zink, 1996). The whole participation process can be a learning experience for ‘designers’ and ‘users’ alike

(e.g. see St-Vincent et al., 1997; Sundin et al., 2004). It can also lead to improvements in processes and employee well-being (e.g. see Lallemand, 2012; Miguez et al., 2012).

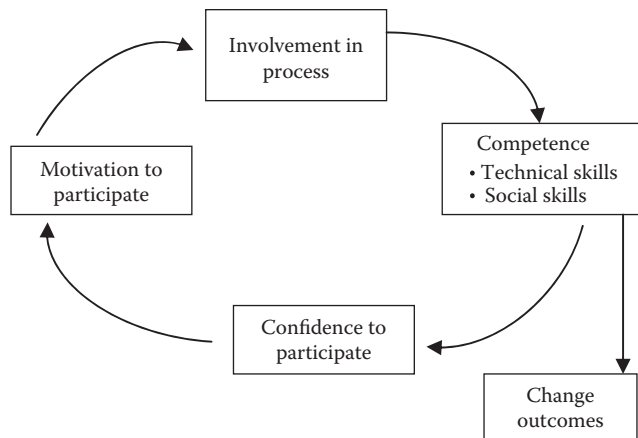
Looking more widely, benefits from the use of participatory approaches may include improved industrial relationships (Lanoie and Tavenas, 1996), opportunities for personal development among those involved and a spread of interest and expertise in ergonomics (Noro, 1991). Participative processes can help embed an ergonomics perspective within the company if people are allowed to make a genuine contribution and influence affairs (Wilson, 1991a,b; St-Vincent et al., 1997).

The relationship between participative processes and the development and gains from working in groups or teams is of great interest. Health and safety and workplace ergonomics investigations can be carried out through work groups. Self-directed team working anyway is often an appropriate mechanism to implement principles of enriched job design, and social support is seen as an effective reducer of strain resulting from stress at work. If work is to be organised on the basis of self-directed teams, then it makes good sense if the groups that develop the ideas for such work organisation and consequent physical workplace requirements will then subsequently become the work teams themselves. In this way the process of change enhances the content. This also meets the sociotechnical principle of comparability (between process and outcome).

Whilst much of the literature on participatory ergonomics is positive, it would be a mistake not to consider some of the difficulties that an ergonomist may encounter when trying to use such an approach within an organisation. It is perhaps understandable that most of the reports in the literature present a positive outcome from the use of participatory methods. After all, organisations that have encountered difficulties from using such an approach may well be resistant to publicise what may be seen as failure. And, even where an initiative has been successful, in the author's experience, participants can often identify aspects of the process or the methods employed that had a negative impact either on themselves or their organisation.

There are a number of reasons why a participatory approach may fail or encounter difficulties. First of all, for whatever reason, some people will just not want to participate or need convincing to participate (e.g. Loch et al., 2010). It is also possible that management might see participation as a threat to their right to manage rather than as an aid (Mumford, 1991). Moreover, participation, as philosophy or process, is neither an easy option nor perhaps always the most appropriate one to choose (Mossink, 1990). A half-hearted or cynical application will reap a harvest of problems, immediately or in the future. Use of participation within the manufacturing industry has certainly often been viewed with suspicion by managers, trade unions and work forces (Garrigou et al., 1995; Loch et al., 2010). Side effects or systemic consequences may also give problems, for instance, increasing the autonomy of assembly work groups may affect the work of the purchasing or maintenance departments; technological advances may reduce the need for personal assistants or information specialists; improvements in one groups' skill levels or learning may make workers in other offices or departments envious or dissatisfied. Related to this last point, the very process of participation may have unexpected systemic effects, in that other groups may wish to be similarly involved but this may not always be feasible or desirable.

Another concern about the use of participation is the quality of outcomes from a process which is based upon group decision making. Some may argue that such a process will only choose the safe option or one that may be viewed as the lowest common denominator. It is therefore important to have established appropriate structures and methods for the participatory group process. Additionally, the development of new workplaces or systems using a participatory approach may take more time and personnel resources and therefore be more costly in the short term. The counter argument to this is, however, that the cost advantages from the introduction of a more efficient and effective solution or design will prove beneficial in the long run and will outweigh extra development costs. Finally, a key issue of work systems change, participative or not, concerns the issue of when the change agent or facilitator should leave the scene. If they leave too late, then true participation may be stifled; if they leave too early, a vacuum may be created in which nobody feels any ownership for the process or change (Wilson, 1995b).



**FIGURE 34.1** Participation cycle. (From Haines, H.M. and Wilson, J.R., *Development of a Framework for Participatory Ergonomics*, HSE Books, Sudbury, Ontario, Canada, 1998.)

From the preceding discussions, it can be seen that participation should not be seen as an easy option. There will be dysfunctional consequences for any participative initiative that is in reality a sham or poorly planned and managed. However, participative initiatives that are well designed and implemented have been shown to lead to a number of benefits both for the organisation and for the participants themselves. The main possible gains for participation are summed up in the participation cycle portrayed in Figure 34.1.

## METHODS USED IN PARTICIPATORY ERGONOMICS

At the heart of any well-designed project is the identification and use of appropriate methods. According to Imada and Nagamachi (1995), the role of participative tools and techniques is ‘to channel participants’ unique knowledge and skills to solve problems and make improvements’ (p. 309). In the literature, a whole range of different tools and methods has been reported as being useful within participatory ergonomics initiatives. However, if we examine this work more closely, it becomes apparent that only some of these tools and methods have been developed or adapted specifically with participatory ergonomics in mind. The majority have been borrowed, either from more traditional E/HF initiatives or from other relevant fields of practice, and then applied within a participative exercise.

One way of thinking about these methods is as a toolkit for participatory ergonomics. Like any toolkit, it contains instruments which are useful for one particular job in the participatory process as well as others that are much more flexible. With this in mind, Table 34.1 shows some of the tools and methods reported in the literature, grouped according to the job for which they might be used within PE initiatives and interventions: preparation and support; problem analysis; creativity stimulation and idea generation; idea/solution generation and development; and evaluation. In reality, of course, there will be a certain degree of overlap between these different groupings.

The participatory tools and methods may differ significantly in the resources they require: not just in terms of time and money but also in the expertise that will be required. There are a number of cases where the analysis of problems is largely tackled by ‘experts’, rather than by workers or end users (e.g. Sundin et al., 2004; Gyi et al., 2012). These experts may well use some of the more complex and technical analytical tools that are available such as QSR International NVivo software or IBM SPSS Statistics – the sophistication lies not so much in their use as in the analysis and interpretation of the results produced. However, as Liker et al. (1989) point out, the ability of organisations to remain self-sufficient goes down as the sophistication of ergonomics analysis tools increases.

**TABLE 34.1**  
**Tools and Methods Used in Participatory Ergonomics**

| Primary Purpose                            | Method or Technique                                                    | References (for examples of where the method or technique has been described or used)                                                                                                                   |
|--------------------------------------------|------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Preparation and support                    | ‘Train the trainers’                                                   | Corlett (1991), Silverstein et al. (1991), Kawakami et al. (2004), Gyi et al. (2012)                                                                                                                    |
|                                            | Breakthrough Thinking®                                                 | Haims and Carayon (1998)                                                                                                                                                                                |
|                                            | Employee Perceptions of Participatory Ergonomics Questionnaire (EPPEQ) | Matthews et al. (2011)                                                                                                                                                                                  |
|                                            | Self-assessment questionnaires                                         | Rosecrance and Cook (2000)                                                                                                                                                                              |
| Problem analysis                           | Posture analysis tool                                                  | Nagamachi and Tanaka (1995), Forsman (2008), Miguez et al. (2012)                                                                                                                                       |
|                                            | ‘Five ergonomic viewpoints’                                            | Imada (1991)                                                                                                                                                                                            |
|                                            | PSIDAR (a video-based system for rating psychosocial aspects)          | Hanse and Forsman (2001)                                                                                                                                                                                |
|                                            | Task analysis, functional task decomposition                           | McNeese et al. (1995), Halpern and Dawson (1997), Rosecrance and Cook (2000), Dos Santos et al. (2011)                                                                                                  |
|                                            | Pareto analysis                                                        | Imada (1991), Hignett et al. (2005)                                                                                                                                                                     |
|                                            | Cause and effect diagram                                               | Imada (1991), Helali (2009)                                                                                                                                                                             |
|                                            | Link analysis                                                          | Imada (1991), Nagamachi (1995)                                                                                                                                                                          |
|                                            | Computer-aided diagrams or drawings                                    | Bengtsson et al. (1997), Berling et al. (1998), Seim and Broberg (2010)                                                                                                                                 |
|                                            | Biomechanical analysis                                                 | Liker et al. (1991), Hess et al. (2004)                                                                                                                                                                 |
|                                            | Videos/photographs                                                     | Algera et al. (1990), Liker et al. (1991), Halpern and Dawson (1997), Rosecrance and Cook (2000), Zalk (2001), Driessen et al. (2008), Forsman (2008), Seim and Broberg (2010), Parimalam et al. (2012) |
|                                            | Brainstorming techniques                                               | Liker et al. (1991), Gjessing et al. (1994), Moore (1994), Rosecrance and Cook (2000), Driessen et al. (2008)                                                                                           |
|                                            | Discussion groups                                                      | Zink (1996), Cohen (1997), Maciel (1998), Hess et al. (2004), Kumar et al. (2005), Driessen et al. (2008)                                                                                               |
|                                            | Prioritised voting or other prioritisation techniques                  | Van der Schaaf and Kragt (1992), Westlander et al. (1995), Driessen et al. (2008)                                                                                                                       |
|                                            | Work-site observation                                                  | Jones (1997), Rosecrance and Cook (2000), Loisel et al. (2001), Parimalam et al. (2012)                                                                                                                 |
|                                            | Interviews, questionnaires                                             | Algera et al. (1990), Van der Schaaf and Kragt (1992), St-Vincent et al. (1998), De Jong and Vink (2002), Dos Santos et al. (2011), Gyi et al. (2012), Parimalam et al. (2012), Guimarães et al. (2014) |
|                                            | Checklists                                                             | Keyserling and Hankins (1994), Laitinen et al. (1997), Moore and Garg (1997), St-Vincent et al. (1998), Rosecrance and Cook (2000), Driessen et al. (2008), Helali (2009)                               |
|                                            | Requirement analysis                                                   | Zink (1996)                                                                                                                                                                                             |
| Creativity stimulation and idea generation | Focus groups                                                           | Caplan (1990), Hess et al. (2004)                                                                                                                                                                       |
|                                            | Design decision groups                                                 | Wilson (1991a)                                                                                                                                                                                          |
|                                            | Problem-solving groups                                                 | Wilson (1995b)                                                                                                                                                                                          |
|                                            | Scenario-driven discussions                                            | Snow et al. (1996), Kawakami et al. (2004)                                                                                                                                                              |

(Continued)

TABLE 34.1 (*Continued*)**Tools and Methods Used in Participatory Ergonomics**

| Primary Purpose                          | Method or Technique                                            | References (for examples of where the method or technique has been described or used)                                                                                                           |
|------------------------------------------|----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Idea/solution generation and development | Layout design games                                            | Seim and Broberg (2010)                                                                                                                                                                         |
|                                          | AKADAM (Advanced Knowledge and Design Acquisition Methodology) | McNeese et al. (1995)                                                                                                                                                                           |
|                                          | Videos/photographs                                             | Algera et al. (1990), Liker et al. (1991), Halpern and Dawson (1997), Kawakami et al. (2004)                                                                                                    |
|                                          | Brainstorming techniques                                       | Liker et al. (1991), Gjessing et al. (1994), Driessen et al. (2008)                                                                                                                             |
|                                          | Scale modelling                                                | Algera et al. (1990), Van der Schaaf and Kragt (1992)                                                                                                                                           |
|                                          | Scenario-driven discussions                                    | Snow et al. (1996), Seim and Broberg (2010)                                                                                                                                                     |
|                                          | Dynamic space manipulation (using visual representations)      | Snow et al. (1996)                                                                                                                                                                              |
|                                          | Role-playing techniques and simulation games                   | Ruohomaki (1995), Seim and Broberg (2010)                                                                                                                                                       |
|                                          | Interviews, questionnaires                                     | Algera et al. (1990), Van der Schaaf and Kragt (1992), St-Vincent et al. (1998), Parimalam et al. (2012)                                                                                        |
|                                          | Virtual reality                                                | Shaikh et al. (2004)                                                                                                                                                                            |
| Evaluation                               | Task analysis                                                  | Kumar et al. (2005)                                                                                                                                                                             |
|                                          | Intervention ideas                                             | Moir and Bucholz (1996)                                                                                                                                                                         |
|                                          | Equipment testing                                              | Van der Schaaf and Kragt (1992), Parimalam et al. (2012)                                                                                                                                        |
|                                          | Biomechanical and postural analysis                            | Kumar et al. (2005), Laing et al. (2005), Driessen et al. (2008), Miguez et al. (2012), Parimalam et al. (2012)                                                                                 |
|                                          | Focus groups                                                   | Caplan (1990), Hess et al. (2004)                                                                                                                                                               |
|                                          | Design decision groups                                         | Wilson (1991a)                                                                                                                                                                                  |
|                                          | Problem-solving groups                                         | Wilson (1995b)                                                                                                                                                                                  |
|                                          | Prioritised voting or other prioritisation techniques          | Van der Schaaf and Kragt (1992), Gjessing et al. (1994), Moore (1994), Westlander et al. (1995)                                                                                                 |
|                                          | Role-playing techniques and simulation games                   | Ruohomaki (1995)                                                                                                                                                                                |
|                                          | Layout modelling and mock-ups                                  | Algera et al. (1990), Kuorinka et al. (1994), St-Vincent et al. (1998), Wilson (2005), Seim and Broberg (2010)                                                                                  |
|                                          | Checklists                                                     | Gjessing et al. (1994), Keyserling and Hankins (1994), Laitinen et al. (1997), Moore and Garg (1997), St-Vincent et al. (1998), Driessen et al. (2008), Helali (2009)                           |
|                                          | Virtual reality                                                | Snow et al. (1996), Shaikh et al. (2004), Määttä (2007)                                                                                                                                         |
|                                          | Sickness absence and injury data                               | Driessen et al. (2008)                                                                                                                                                                          |
|                                          | Interviews, questionnaires                                     | Algera et al. (1990), Van der Schaaf and Kragt (1992), St-Vincent et al. (1998), De Jong and Vink (2002), Laing et al. (2005), Gyi et al. (2012), Miguez et al. (2012), Guimarães et al. (2014) |
|                                          | Research diaries                                               | Gyi et al. (2012)                                                                                                                                                                               |
|                                          | Team formation and building                                    | Caccamise (1995), Kawakami et al. (2004)                                                                                                                                                        |
|                                          | Team training                                                  | Gjessing et al. (1994), Garmer et al. (1995), Loisel et al. (2001), Driessen et al. (2008)                                                                                                      |

As one of the aims of participatory ergonomics may be to free institutions from reliance on the outside expert, the use of highly complex tools may prove self-defeating.

## DESIGN DECISION GROUPS

Design decision groups (DDGs) combine a number of the techniques mentioned earlier in this section – word maps, round-robin questionnaires, layout modelling and mock-ups – into a systematic approach traditionally for workplace layout and design (Wilson, 1991a). This concept is derived from the work of Dennis O'Brien (of the UK Home Office) into methods for running creative meetings and for involving users in systems design (e.g. O'Brien, 1981). O'Brien adapted theories and techniques from market research and the literature on creativity and innovation, in particular using non-directive and group discussions, developing 'thinking tools' and encouraging participants to talk about, write down or draw, attitudes experiences, visions or opportunities (see also Caplan, 1990). O'Brien described a number of techniques or tools within what he called shared experience events (SEEs) and among these are structuring of contributions from sub-groups whilst others listen in silence, only later criticising; use of drawing as far as possible, which enhances concept development, communication of ideas and also helps in structuring the process; and very careful pre-planning and preparation. A good selection of case studies involving similar techniques/methods to support design decisions is found in Langford and McDonagh (2003). The DDG process proposed by Wilson (1991a) is summarised in Figure 34.2, and further description and illustration of the stages appear as a case study in the appendix to this chapter.

### Extension of the DDG Process

One criticism often made of DDG and similar techniques is that they are usually only employed with white-collar or technical staff, but there are examples where this is not the case, for example Wilson (1995b), Dos Santos et al. (2011) and Seim and Broberg (2010). The adapted DDG reported by Wilson (1995b) concerns crane drivers operating from a control room in an incinerator plant. Early observations identified that they had severe problems related to viewing the task, the design of a deadman's handle and joystick controls, seating and postural inadequacies and general environmental impacts. A form of DDG technique was adapted for use with the drivers; critically this involved them in sourcing and costing their preferred design solutions within a budget set by management. The process is illustrated in Table 34.2. It can be seen that the final stage is that of continual improvement – the embedding in of participation as *the* way forward in ergonomics problem assessment and solution.

The use of a participative process in this study probably did not lead to the identification of any different problems and causes than would have been the case with a traditional ergonomics approach. However, the perspective on these problems and the priorities given them was different, and this guided subsequent improvements in directions which best met the needs and interests of the workers. The solutions which were eventually agreed were, at best, satisfying; that is it could not be argued that they were the best that could have been developed but they did entail a satisfactory answer to the problems. Since there is always the question of what is 'best' in ergonomics anyway, and since the process of continual change has led to further improvements, the outcome was considered to be successful. The change agents withdrew from the process as soon as possible, with the consequence that the proposed redesigns were only partially implemented and therefore from their viewpoint there were still some ergonomics difficulties. However, this seems to be outweighed by finding not only that the changes made were completely accepted, widely welcomed and gave rise to a dramatic fall in complaints and postural problems, but that the group has continued to discuss and implement further workplace and work improvements, demonstrating that an ergonomics perspective has been embedded for that group of workers to at least a limited extent.



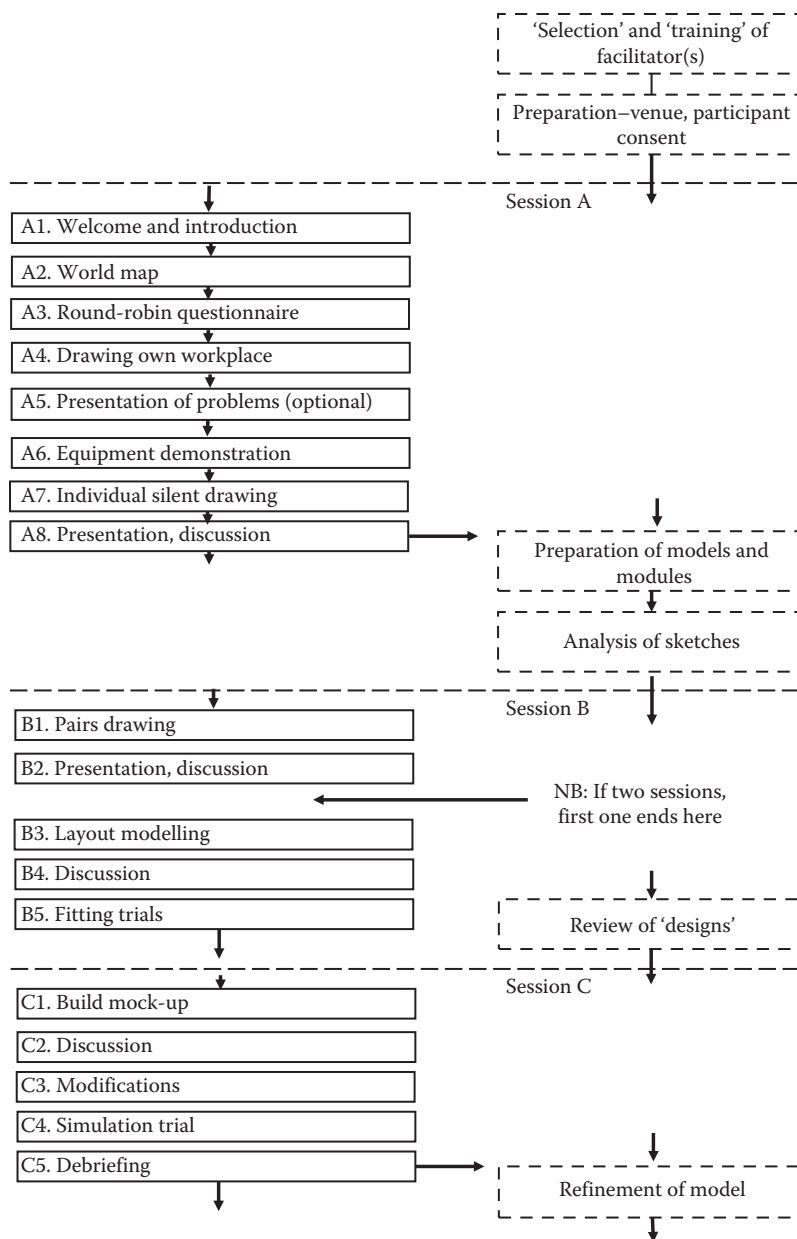


FIGURE 34.2 Summary of the design decision group process.

## PARTICIPATORY ERGONOMICS FRAMEWORK

The fact that PE can operate at a number of different levels and encompasses a range of different approaches can make the study of such initiatives quite a challenge. In part to address this issue, a framework for the guidance and review of participatory ergonomics projects was developed by Haines et al. (2002). The participatory ergonomics framework (PEF) describes a number of different dimensions across which participatory initiatives may vary. Each dimension has two or more associated categories that further define this feature of the initiative (see Table 34.3).

When using the framework, there are two important points to consider. The first is to identify the number of different participatory groups that may exist within an organisation. For example,

**TABLE 34.2**  
**Stages of the Procedure**

**DOC**

|          |                                                                                                                                                      |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| Stage 1: | Familiarisation: of investigator with the work and workplace and of the participants with ergonomics and the aims of the PSG                         |
| Stage 2: | Field visits: by participants to 'similar' sites                                                                                                     |
| Stage 3: | Design decision groups (A): elements of session A of the DDG process                                                                                 |
| Stage 4: | Design decision groups (B): elements of session B of the DDG process                                                                                 |
| Stage 5: | Lighting simulation: using screens, coverings, etc., in the control room, to reduce reflection problems, etc.                                        |
| Stage 6: | Sourcing and costing of solutions: crane drivers finding products or adaptation, which best meet their preferred solutions, whilst staying in budget |
| Stage 7: | Continual improvement: after the facilitators have left                                                                                              |

*Sources:* Wilson, J.R., Ergonomics and participation, in Wilson, J.R. and Corlett, E.N., eds., *Evaluation of Human Work: A Practical Ergonomics Methodology*, 2nd and revised edition, Taylor & Francis Group, London, U.K., pp. 1071–1096, 1995a; Wilson, J.R., *Int. J. Ind. Ergon.*, 15, 329, 1995b.

**TABLE 34.3**  
**Participatory Ergonomics Framework (PEF)**

| <b>Dimension</b>              | <b>Categories</b>                                                                                                                                                   |
|-------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Permanence                    | Ongoing – temporary                                                                                                                                                 |
| Involvement                   | Full direct participation – direct representative participation – delegated participation                                                                           |
| Level of influence            | Group of organisations – entire organisation – department – work group/team                                                                                         |
| Decision making               | Group delegation – group consultation – individual consultation                                                                                                     |
| Mix of participants           | Operators – line management – senior management – internal specialist/technical staff – union – external advisor – supplier/purchaser – cross industry organisation |
| Requirement                   | Compulsory – voluntary                                                                                                                                              |
| Focus                         | Physical design/specification of equipment/workplaces/work tasks – design of job teams or work organisation – formulation of policies or strategies                 |
| Brief                         | Problems identification – solution development – implementation of change – set-up/structure process – monitor/oversee process                                      |
| Role of ergonomics specialist | Initiates and guides process – acts as expert – trains participants – available for consultation – not involved                                                     |

*Source:* Haines, H.M. et al., *Ergonomics*, 45(4), 309, 2002.

a participatory initiative may involve employee teams focussed on workplace improvement, health and safety committees and structured organisational ergonomics teams. In such cases, it is suggested that each of these groups should be evaluated separately. Secondly, within each of these groups it is important to clarify whose perspective is being considered when looking at the different categories in each of the dimensions. This is because the evaluation may well differ for each stakeholder. For example, whilst operators may volunteer to take part in the participatory group, if there is only one line manager then they may be required to participate.

In addition to allowing different participatory interventions to be compared or evaluated, the PEF may also have a role in guiding the planning of a participatory project (Haines et al., 2002). It is suggested that the various dimensions and their categories can be used to structure discussions

during the planning phase of such a project. As a result, the process to be used, the boundaries of the group, definitions of participants' roles, etc. may be agreed at the outset. Such considerations should ensure that as much attention is given to the process of participation as to the content of the intervention. In setting up a PE project, Gyi et al. (2012) also highlight the importance of the 'lead time' and securing management 'permissions' – this period can be lengthy and therefore should be considered part of the PE process itself. In addition, procurement should ensure the long-term sustainability of any interventions and consider ways of measuring impact. The PE process should also be flexible to adapt to changes in organisational structures, internal politics, key personnel and global influences.

### REQUIREMENTS FOR PARTICIPATORY INITIATIVES

One remaining question for those considering the use of PE might be 'what do we need for a successful participatory initiative?' This issue has been considered by a number of authors (e.g. De Looze et al., 2001; Vink et al., 2006; Loch et al., 2010; Gyi et al., 2012), and the following were identified as particularly important:

- Commitment to the project at all levels of the organisation, that is management and worker involvement
- A strong reason to take a participative approach
- A champion to support and/or facilitate the process
- Establishing a steering group with clear responsibilities
- Clear definition of the actors and their roles – who will be involved?
- Structures to support the process – how will the participation be managed?
- Keeping the project simple – well defined and well targeted – a step-by-step approach
- Having an understanding of the impact, that is the benefits in both quantitative and non-quantitative measures
- The need for training in general ergonomics principles

In addition, the following may also be helpful (St-Vincent et al., 1997; Haines and Wilson, 1998; Ingelgard and Norrgren, 2001):

- Appropriate levels of knowledge for all participants
- Previous good experience of participation
- Trade union involvement
- The need for training in teamwork and interaction skills

The role of the ergonomist in a participatory ergonomics initiative was also highlighted as requiring careful thought and definition. Their input may vary from one of 'expert' to group facilitator, either between different participatory projects or within a single project (Wilson, 1995b; Haims and Carayon, 1998). It is interesting that in the review of PE interventions by Van Eerd et al. (2010), an expert, often an ergonomist was the facilitator for the process. Finally, the issue of sufficient resources for the project is a vital one. Senior management agreement for the financial, personnel and time resources required is essential for setting up and running a participatory project. The preparation and time taken to secure such 'permissions' and effect change at management level should also not be underestimated and is an important part of the PE process (Gyi et al., 2012). Finally, mechanisms to evaluate the impact of participatory initiatives are also increasingly important. In summary, some of the requirements for a successful PE initiative are shown in Table 34.4 (Wilson and Haines, 1997).

**TABLE 34.4**  
**Requirements for Participatory Ergonomics**

|                           |                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Climate for participation | Avoid times of uncertainty, hostility, crisis<br>Better in open, non-hierarchical, non-centralised organisations<br>Support if group work and enriched jobs in place already<br>Devolve ergonomics: foundation/spread/embed                                                                                                                                                                          |
| Support and resources     | Support throughout the organisation, and trade unions<br>Key decision-maker commitment<br>Expect resistance – rational and irrational<br>Allocate sufficient finance and personnel<br>No unreasonable time constraints<br>Set initial flexible goals and criteria                                                                                                                                    |
| Facilitator               | Is not necessarily project ‘owner’<br>Sensitive yet robust and decisive<br>Unbiased yet knowledgeable<br>Flexible and consistent<br>Respected by participants and organisation<br>Knows when to leave the process                                                                                                                                                                                    |
| Set-up                    | Timing and starting point are critical<br>Process has to have initiation and evolution<br>Interdependence of motivation, ability and confidence of participants<br>Participation to be voluntary where possible<br>Potential benefits should be promoted during set-up                                                                                                                               |
| Processes and methods     | Participants can be self-selecting; otherwise use volunteers, elections or selection<br>Size of group: 6–10<br>Need facilitator, chairperson, resources assistant<br>Methods to be well costed, cost-effective and flexible<br>Process to be iterative, relaxed and allow compromises<br>Methods to promote creativity, learning, confidence and decision making<br>Include solution-costing methods |
| Outcomes                  | Need to have direction without being overtly directed<br>Testable solutions or plans<br>Continual improvements<br>Spread of expertise in the organisation<br>Process to become internal and ordinary, not external and extraordinary                                                                                                                                                                 |

*Sources:* Adapted from Wilson, J.R., *Int. J. Ind. Ergon.*, 15, 329, 1995b; Wilson, J.R. and Haines, H.M., Participatory ergonomics, in Salvendy, G., ed., *Handbook of Human Factors and Ergonomics*, 2nd edn., John Wiley & Sons, Chichester, U.K., 1997, pp. 490–513.

## CONCLUSIONS

There is a strong argument for a participative approach to ergonomics: both participation at an individual level – whereby workers are involved in work/workplace assessment and in providing solutions to any problems identified; and also participation at an organisational level – whereby engineering, production, medical and other staff develop their own programmes for ergonomics awareness and action. There is a danger that participation may be seen as an easy option when, in reality, it requires careful planning and implementation. It is important to identify why a participatory ergonomics approach should be used and follow good practice. The wide range of case studies in the literature provide examples of projects within different industries and opportunities for lessons to be learnt from others’ experiences. Of course, there are limits to PE: there are certain

problems and solutions that will always require professional (often outside) ergonomics expertise. At least a part of the training given should be aimed to allow companies to understand not just what they can do internally but also what they should not attempt.

Finally, efforts to enable the widespread and successful implementation of ergonomics at work have multiplied in recent years; strenuous efforts are being made to take ergonomics out of the pages of text books and make it more applicable in practical use. PE has played a part in this – we cannot just rely upon ‘ergonomics experts’ – there are not enough of these, it will not be cost-effective for most smaller organisations and is, anyway, not the right way to approach the challenge. The challenge is to

- Motivate people to become involved in ergonomics within their own organisation as part of a process that has meaning and influence
- Give them tools and techniques and related training to make ergonomics interventions, ensuring that training comes at the right time for all relevant groups
- Ensure they have enough understanding to know when and how to use these tools and techniques, *and when they cannot and must call in specialist assistance*

In fact, at a more general level, there is little in an ergonomics and a participative approach which is not a part of what should be good practice within management and organisational culture.

## APPENDIX: PARTICIPATORY ERGONOMICS CASE STUDY

This case study has been adapted from Wilson (1991a) – although from some time ago, it is still a useful example, containing good practice ideas on how to conduct DDGs and/or focus groups in participatory ergonomics. It will be of particular interest to researchers and practitioners new to this type of approach.

## DESIGN DECISION GROUPS PROCEDURE

### SCENE SETTING

In order to enable (not force) people to participate in meetings, exercises, investigations, etc., certain general principles can be observed. They should feel that there is some purpose to participating, that outcomes will be of benefit to themselves and/or others, confident enough to ask questions, and that they have enough information and knowledge to make a useful input. One important but tricky requirement is that the group must not be overtly directed or feel that this is the case; however, the group must also have direction and be purposeful; disruptive or domineering members must be subtly reined in and not allowed to dominate meetings.

These requirements can be achieved in several ways. In this case study of retail checkouts, holding meetings at an independent site (a laboratory) and emphasising this independence and the confidentiality of the details of the process were important. Meetings were facilitated rather than run, generally with one investigator acting as introducer, prompter, interpreter and general resource, with a second investigator on hand as technical resource, including making and serving refreshments, distributing or showing graphical and visual aids, supplying materials and so on. The second investigator was also responsible for recording the proceedings, using video and photography (as unobtrusively as possible), and for note taking. All sessions were allowed to run in a very relaxed manner, with everybody being gently encouraged to contribute, and with a very flexible agenda. At the same time, the investigators were well aware of the tasks or exercises which had to be completed, and information which had to be derived, within particular time-scales. Although guidelines for doing this can be followed and can help, it is not surprising that it also takes skill to facilitate groups well; not everyone has the ability to communicate and subtly direct, without dictating or dominating.

In this case study, all potential participants were contacted at their place of work and asked if they would take part in confidential discussions about the design of their workplace. They were told how much they would be paid, and the dates and times they would be needed and that they would be collected and returned by taxi. Payment, which compared very favourably with their normal wages, was made in one lump sum at the end of all sessions. The procedure for each session is now described.

## SESSION 1

### INTRODUCTION

This included a welcome, introductions, an explanation of who the design meeting organisers were and a general overview of what was expected of the design sessions. Refreshments were served and every effort was made to establish a comfortable, relaxed environment.

### WORD MAP

'Word maps' or 'visual maps' were constructed by every participant, volunteering words, which they felt were connected with the particular design topic for a 5–10 min period. This technique defined the total problem space and helped produce the dynamic agenda. In this case study, participants were asked to make an inventory of objects and equipment required at their workplace (see Figure 34.3). This acted as an 'ice-breaking' technique and helped break down any inhibitions that quieter members of the group might have about expressing their ideas in front of their colleagues. Participation was encouraged by going 'round the table' several times.

### ROUND-ROBIN QUESTIONNAIRE

Participants were presented with a series of very simple open-ended questions such as: 'A good work counter is...?' or 'Problems at my workplace are...?' Each question was printed on a separate sheet, and these were passed around all participants with each attempting to complete the sentence with a different ending to that chosen by the others. This method is preferred to the alternative of each participant filling out an individual questionnaire for two reasons. Firstly, it makes the exercise more productive, in that each participant has to think of a different ending from those given by the others, thus providing more information and reducing duplication. Secondly, it can act as a stimulus. If, for example, participants had

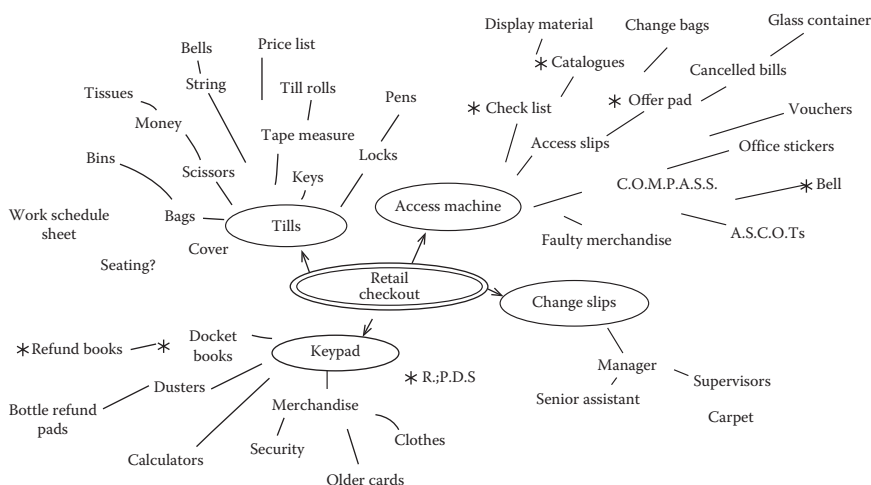


FIGURE 34.3 Part of a simple word map for a retail checkout.



missed the point of the exercise, or were unable to think of any ideas for a certain sentence, then seeing what had been written by others may trigger a line of thought that could otherwise have lain dormant.

### Drawing of Their Own Workplace

As a further ice-breaking exercise and in preparation for the drawing exercises to come, whilst also providing investigators with extra information, participants were asked to draw their existing workplace, indicating its good and bad points.

### Slide Presentation of Ergonomics Problems

Slides were selected from associated fieldwork on retail checkouts (in this case), covering a range of topics; different types of layout, different types of equipment, etc., and most importantly, showing how equipment has been poorly installed in some work environments. The slides emphasised features such as poor seat height, lack of leg room, poor work-surface height, lack of work space and poor positioning of equipment and displays.

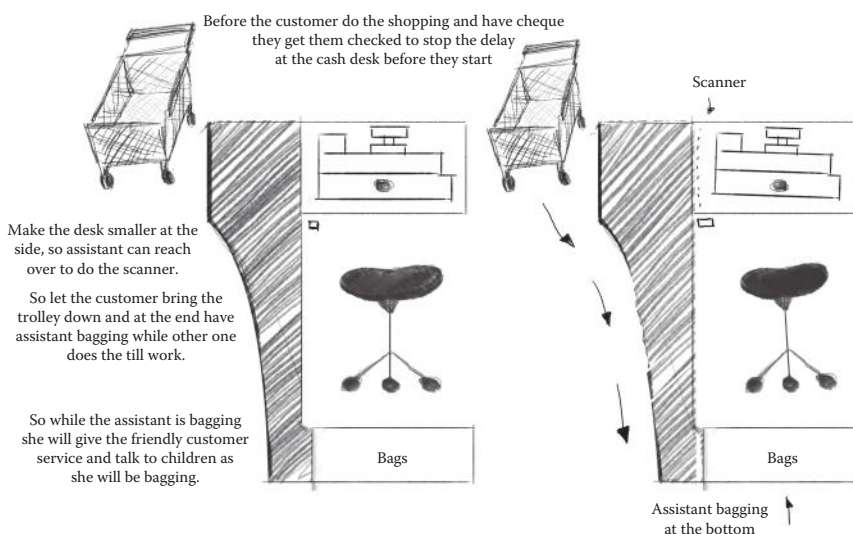
This stage is intended to provide a logical progression from the round-robin questionnaire. At that stage, the participants were asked to think about what makes good and poor designs, but only had their own experience to draw upon. The slide presentation was intended to make participants aware of systems and layouts that they may have not encountered before, and also to emphasise problems they may not have considered. Slides of a variety of different workplaces and systems also emphasised the point that there was an opportunity to design something entirely new, and that suggested designs should be unimpeded by a narrow range of experience.

### Equipment Demonstration

When systems or equipment are to be installed which may be new to any or all participants, it is advisable to demonstrate these and their operation. Actual examples should be used where possible; if not visual representations (e.g. photographs/diagrams/videos/models) can be used.

### Individual Silent Drawing

Each participant was asked to produce sketches on A4 paper of both plan and perspective views of their ideas for a good new workplace, incorporating any new systems. These should detail proposed layouts for people, work counters, modules and equipment (see Figure 34.4 for



**FIGURE 34.4** A reproduction of an annotated sketch of a proposed workspace.

an example). At this stage, only very general ideas need to be produced, and participants were discouraged from becoming too involved in the detail (e.g. precise measurements of work surfaces or exact locations of documents). To call this stage a 'silent drawing exercise' is perhaps overstating the point, but it is beneficial to discourage communication in the hope that a variety of different ideas and designs will be produced for use in the later discussion stage. Each individual is encouraged to try out several rough ideas before producing a final sketch suitable for presentation. Different coloured pens were available for coding various items of equipment or furniture.

### **Individual Presentation and Discussion**

Each participant was asked to present their final sketch to the rest of the group and to explain the reasoning behind their design. The designs were discussed and gently (the role of the facilitator) criticised.

After each of the sketches had been discussed, many problems were highlighted. The role of the facilitator is then to encourage a general discussion, concentrating on one or two particular designs/ideas and bringing up any points that may not have been previously considered, probably referring back to issues arising from the word-map and round-robin stages.

## **SESSION 2**

### **DRAWING IN PAIRS**

At the beginning of the session, the facilitator briefly ran through some of the points arising from the sketches made in session 1. The group was then divided into pairs or threes, according to the similarity of their ideas, and asked to produce a larger sketch on A3 paper, of their plan for the workplace. This time more emphasis is placed on the detail, for example precise positioning of equipment, and seating and furniture design.

### **Presentation and Discussion**

Each drawing was presented to the rest of the group by its authors. Discussion at this point is geared towards choosing the 'best' ideas, but no firm decisions need to be reached as each group gets a further chance to demonstrate their design using simple physical models.

### **Simple Layout Modelling**

Participants were shown empty spaces of approximately the same size available at the future workplace site. Very strong corrugated cardboard models of equipment were provided, in addition to a supply of mock-up counters, tables, desks, consoles and other modules. A variety of actual seats, tables and benches were also available, as well as string, flat card, modelling knives and heavy tape. There are several compelling advantages of these models – they are cheap, portable, easy to re-position and participants were not afraid of breakage or changing the design – building model workstations became a fast changing and dynamic exercise. Extra pieces of equipment could also be found as necessary and new module sizes and pieces for the new work-station could be very quickly and easily created.

Participants, still in subgroups of two or three, were then asked to build a 3D full-size model of their idea of their optimum workplace design.

### **Discussion**

Discussion involving all participants is firmly geared toward eliminating any inferior design(s) and forming a consensus of opinion as to the optimal layout. This is often an amalgam of the various ideas represented by different mock-ups. Limited 'walkthroughs', testing aspects of the workplace may be undertaken at this stage.

### **Fitting Trials**

On reaching this stage, a general workplace layout and positioning of all the equipment should largely have been agreed. The next crucial factor to be tackled was that of work-surface heights and depths. These were determined through fitting trials involving all participants, with the aid of adjustable metal staging holding the cardboard models. In addition, the facilitator must ensure, at this time or later, that anthropometric characteristics of other potential users are taken into account.

## **SESSION 3**

### **BUILDING A MOCK-UP**

On returning for the third session, participants built a complete mock-up of the preferred workstation, integrating ideas on the placement and juxtaposition of equipment with work-surface height and depth data. If costs permitted, a second mock-up could have been built if there was internal disagreement and no consensus on a final design.

### **Discussion**

The discussion entailed detailed consideration of the operational requirements at each particular workstation; if two designs had been modelled, each sub-group must defend its ideas against those of the other group.

### **Modifications**

As a result of the discussion, any modifications to the general structure of the workstations were made. If more than one design is still being considered, the best alternative is generally selected at this stage.

### **Simulation Trial and Assessment**

The trials are aimed at simulating as many as possible of the scenarios, operations and tasks which are or might be undertaken in actual use. Participants took turns in playing the roles of staff and customers, including children. Final adjustments to the workplace model are made at this point.

### **Debriefing**

Debriefing takes place after each session to fully understand what is being communicated, but the final debriefing is the most important. Participants are encouraged to discuss the value of the exercise and the perceived quality of their designs. Refinements were made at the end of each stage, with the investigators accounting for constraints, equipment or tasks which may have been omitted (deliberately or accidentally) in the sessions, or which may have come to attention later.

## **OUTCOMES**

The outcomes of this case study were varied. At the very least the investigators managed to find out from the people who should know most about it, the problems or benefits associated with existing systems, the potential problems with the new system and a number of concepts to be fed into the development process. At best, one or more feasible design alternatives for the whole workstation were provided; this was an invaluable outcome, regardless of whether these designs were 'new' to the investigators or were combinations of ideas already developed. The process and setting also allow flexible and iterative testing of design concepts, whether derived beforehand or during the process.

It was noted by the investigators that the creative, yet relaxed, atmosphere of the DDGs and the resultant concentration of minds seemed to encourage the expression of opinion or facts about other aspects of work redesign. Much was also learned about training and support, feedback

and other communication, autonomy, role clarity and other job-content and organisation issues. This allowed a more informed analysis of current situations, and to make better suggested changes for improvement. It is suggested that the spin-off benefits will be considerable when such an exercise is conducted at a local level with people whose own technology, workplace and jobs are changed. Awareness of what should be done to improve jobs and workplaces and of the criteria which should be applied, and the confidence to promote their own ideas can all help to create a more committed workforce, as well as healthier, safer, more effective and more satisfying jobs. The use of this type of approach can thus give rise to systematic gains through the very application of the process.

## REFERENCES

- Algera, J.A., Reitsma, W.D., Scholtens, S., Vriens, A.A.C. and Wijnen, C.J.D. (1990). Ingredients of ergonomic intervention: How to get ergonomics applied. *Ergonomics*, **33**(5), 557–578.
- Bengtsson, P., Johansson, C. and Akselsson, K.R. (1997). Planning working environment and production by using paper drawings and computer animation. *Ergonomics*, **40**(3), 334–347.
- Berling, C., Blome, M., Johansson, C., Odenrick, P. and Rassner, F. (1998). Methods for introducing improvements at work. In: P. Vink, E. Koningsveld and S. Dhondt, eds., *Human Factors in Organisational Design and Management – VI*. London, U.K.: Elsevier Science, pp. 555–565.
- Boel, M., Daniellou, F., Desmores, E. and Teiger, C. (1985). Real work analysis and workers' involvement. In: I. Brown, R. Goldsmith and M. Sinclair, eds., *Ergonomics International 85, Proceedings of the Ninth Congress of the International Ergonomics Association*. London, U.K.: Taylor & Francis Group, pp. 235–237.
- British Standards Institution. (2007). BS OHSAS18001: 2007. Occupational Health and Safety Management Systems – requirements. London, U.K.: BSI.
- Caccamise, D.J. (1995). Implementation of a team approach to nuclear criticality safety: The use of participatory methods in macroergonomics. *International Journal of Industrial Ergonomics*, **15**, 397–409.
- Caplan, S. (1990). Using focus group methodology for ergonomic design. *Ergonomics*, **33**, 527–533.
- Cohen, R. (1997). Ergonomics program development: Prevention in the workplace. *American Industrial Hygiene Association Journal*, **58**, 145–149.
- Corlett, E.N. (1991). Ergonomics fieldwork: An action programme and some methods. In: M. Kumashiro and E.D. Megaw, eds., *Towards Human Work: Solutions to Problems in Occupational Health and Safety*. London, U.K.: Taylor & Francis Group, pp. 179–185.
- Cotton, J.L. (1993). *Employee Involvement: Methods for Improving Performance and Work Attitudes*. Newbury Park, CA: Sage Publications.
- De Jong, A.M. and Vink P. (2002). Participatory ergonomics applied in installation work. *Applied Ergonomics*, **33**, 439–448.
- De Looze, M., van Rhijn, G., Tuinzaad, B. and van Deursen, J. (2000). A participatory and integrative approach to increase productivity and comfort in assembly. *Ergonomics for the New Millennium. Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Meeting of the Human Factors and Ergonomics Society*, San Diego, CA, vol. 5, pp. 142–145.
- De Looze, M.P., Urlings, I.J.M., Vink, P., Van Rhijn, J.W., Miedema, M.C., Bronkhorst, R.E. and Van der Grinten, M.P. (2001). Towards successful physical stress reducing products: an evaluation of seven cases. *Applied Ergonomics*, **32**(5), 525–534.
- Dos Santos, I.J.A.L., Farias, M.S., Monteiro, B.G., Falcão, M.A. and Marcelino, F.D. (2011). Using participatory ergonomics to improve nuclear equipment design. *Journal of Loss Prevention in the Process Industries*, **24**, 594–600.
- Driessen, M.T., Anema, J.R., Proper, K.I., Bongers, P.M. and van der Beek, A.J. (2008). Stay@Work: Participatory ergonomics to prevent low back and neck pain among workers: Design of a randomised controlled trial to evaluate the (cost-) effectiveness. *BMC Musculoskeletal Disorders*, **9**, 145.
- Eason, K. (1995). User-centred design: For users or by users? *Ergonomics*, **38**, 1667–1673.
- Forsman, M. (2008). Participatory ergonomics—supporting tools. *Proceedings of the 13th International Conference on Productivity and Quality Research, ICPQR 2008 Conference*, Oulu, Finland, 25–27 June 2008, pp. 195–204.
- Garmer, K., Dahlman, S. and Sperling, L. (1995). Ergonomic development work: Co-education as a support for user participation at a car assembly plant. A case study. *Applied Ergonomics*, **26**(6), 417–423.

- Garrigou, A., Daniellou, F., Carballeda, G. and Ruaud, S. (1995). Activity analysis in participatory design and analysis of participatory design activity. *International Journal of Industrial Ergonomics, Special Issue: 'Participatory Ergonomics'*, **15**(5), 311–329.
- Gjessing, C.C., Schoenborn, T.F. and Cohen, A. (1994). *Participatory Ergonomics Interventions in Meatpacking Plants*. Cincinnati, OH, USA: DHSS (NIOSH) Publication.
- Glew, D.J., O'Leary-Kelly, A.M., Griffin, R.W. and Van Fleet, D.D. (1995). Participation in organisations: A preview of the issues and proposed framework for future analysis. *Journal of Management*, **21**(33), 395–421.
- Guimarães, L.B.M., Ribeiro, J.L.D., Renner, J.S. and de Oliveira, P.A.B. (2014). Worker evaluation of a macro-ergonomic intervention in a Brazilian footwear company. *Applied Ergonomics*, **45**(4), 923–935.
- Gyi, D., Sang, K. and Haslam, C. (2012). Participatory ergonomics: Co-developing interventions to reduce the risk of musculoskeletal symptoms in business drivers. *Ergonomics*, **56**(1), 45–58.
- Haims, M.C. and Carayon, P. (1998). Theory and practice for the implementation of 'in-house' continuous improvement participatory ergonomics programs. *Applied Ergonomics*, **29**(6), 461–472.
- Haines, H.M. and Wilson, J.R. (1998). *Development of a Framework for Participatory Ergonomics*. Sudbury, Ontario, Canada: HSE Books.
- Haines, H.M., Wilson, J.R., Vink, P. and Koningsveld, E. (2002). Validating a framework for participatory ergonomics (the PEF). *Ergonomics*, **45**(4), 309–327.
- Halpern, C. and Dawson, K. (1997). Design and implementation of a participatory ergonomics program for machine sewing tasks. *International Journal of Industrial Ergonomics*, **20**, 429–440.
- Hanse, J. and Forsman, M. (2001). Identification and analysis of unsatisfactory psychosocial work situations: A participatory approach employing video-computer interaction. *Applied Ergonomics*, **32**, 23–29.
- Haslam, C., Clemes, S., Crawford, J., Gibb, A., Gyi, D., Haslam, R., McDermott, H., Morgan, K. and Nicolle, C. (2013). NDA findings 22: Working late. New dynamics of ageing research programme, <http://www.workinglate.org/nda-findings-22/> (accessed 21 March 2014).
- Health and Safety Executive. (2002). *Upper Limb Disorders in the Workplace*. Sudbury, Ontario, Canada: HSE Books.
- Health and Safety Executive. (2013). *Managing for Health and Safety*. Sudbury, Ontario, Canada: HSE Books.
- Helali, F. (2009). Using ergonomics checkpoints to support a participatory ergonomics intervention in an industrially developing country (IDC) – a case study. *International Journal of Occupational Safety and Ergonomics*, **15**(3), 325–337.
- Hess, J.A., Hecker, S., Weinstein, M. and Lunger, M. (2004). A participatory ergonomics intervention to reduce risk factors for low-back disorders in concrete laborers. *Applied Ergonomics*, **35**(5), 427–441.
- Hignett, S., Wilson, J.R. and Morris, W. (2005). Finding ergonomic solutions – participatory approaches. *Occupational Medicine*, **55**, 200–207.
- Imada, A.S. (1991). The rationale and tools of participatory ergonomics. In: K. Noro and A.S. Imada, eds., *Participatory Ergonomics*. London, U.K.: Taylor & Francis Group, pp. 30–51.
- Imada, A.S. and Nagamachi, M. (1995). Introduction to a special issue on participatory ergonomics. *International Journal of Industrial Ergonomics*, **15**, 309–310.
- Ingelgard, A. and Norrgren, F. (2001). Effects of change strategy and top-management involvement on quality of working life and economic results. *International Journal of Industrial Ergonomics*, **27**(1), 93–105.
- Jones, R.J. (1997). Corporate ergonomics program of a large poultry processor. *American Industrial Hygiene Association Journal*, **58**, 132–137.
- Kardborn, A. (1998). Inter-organisational participation and user focus in a large-scale product development programme: The Swedish hand tool project. *International Journal of Industrial Ergonomics*, **21**(5), 369–381.
- Kawakami, T., Kogi, K., Toyama, N. and Yoshikawa, T. (2004). Participatory approaches to improving safety and health under trade union initiative. *Industrial Health*, **42**, 196–206.
- Keyserling, W.M. and Hankins, S.E. (1994). Effectiveness of plant-based committees in recognizing and controlling ergonomic risk factors associated with musculoskeletal problems in the automotive industry. *Rehabilitation*, **3**, 346–348.
- Kumar, R., Chaikumarn, M. and Lundberg, J. (2005). Participatory ergonomics and an evaluation of a low-cost improvement effect on cleaners' working posture. *International Journal of Occupational Safety and Ergonomics*, **11**(2), 203–210.
- Kuorinka, I. (1997). Tools and means of implementing participatory ergonomics. *International Journal of Industrial Ergonomics*, **19**, 267–270.
- Kuorinka, I., Cote, M.-M., Baril, R., Geoffrion, R., Giguere, D., Dalzell, M.A. and Larue, C. (1994). Participation in workplace design with reference to low back pain: A case for improvement of the police patrol car. *Ergonomics*, **37**, 1131–1136.



- Laing, A., Frazer, M., Cole, D., Kerr, M., Wells, R. and Norman, R. (2005). Study of the effectiveness of a participatory ergonomics intervention in reducing worker pain severity through physical exposure pathways. *Ergonomics*, **48**(2), 150–170.
- Laitinen, H., Saari, J. and Kuusela, J. (1997). Initiating an innovative change process for improved working conditions and ergonomics with participation and performance feedback: A case study in an engineering workshop. *International Journal of Industrial Ergonomics*, **19**, 299–305.
- Lallemant, C. (2012). Contributions of participatory ergonomics to the improvement of safety culture in an industrial context. *Work*, **41**, 3284–3290.
- Langford, J. and McDonagh, D. (2003). *Focus Groups: Supporting Effective Product Development*. London, U.K.: Taylor & Francis Group.
- Lanoie, P. and Tavenas, S. (1996). Costs and benefits of preventing workplace accidents: The case of participatory ergonomics. *Safety Science*, **24**(3), 181–196.
- Liker, J.K., Joseph, B.S. and Ulin, S.S. (1991). Participatory ergonomics in two US automotive plants. In: K. Noro and A. Imada, eds., *Participatory Ergonomics*. London, U.K.: Taylor & Francis Group, pp. 97–139.
- Liker, J.K., Nagamachi, M. and Lifshitz, Y.R. (1989). A comparative analysis of participatory ergonomics programs in U.S. and Japan manufacturing plants. *International Journal of Industrial Ergonomics*, **3**, 185–189.
- Loch, C.H., Sting, F.J., Bauer, N. and Mauermann, H. (2010). How BMW is defusing the demographic time bomb. *Harvard Business Review*, **88**(3), 99–102.
- Loisel, P., Gosselin, L., Durand, P., Lemaire, J., Poitras, S. and Abenhaim, L. (2001). Implementation of a participatory ergonomics program in the rehabilitation of workers suffering from subacute back pain. *Applied Ergonomics*, **32**, 53–60.
- Määttä, T.J. (2007). Virtual environments in machinery safety analysis and participatory ergonomics. *Human Factors and Ergonomics in Manufacturing*, **17**(5), 35–443.
- Maciel, R. (1998). Participatory ergonomics and organisational change. *International Journal of Industrial Ergonomics*, **22**(4–5), 319–325.
- Matthews, R.A., Gallus, J.A. and Henning, R.A. (2011). Participatory ergonomics: Development of an employee assessment questionnaire. *Accident Analysis and Prevention*, **43**(1), 360–369.
- McNeese, M.D., Zaff, B.S., Citera, M., Brown, C.E. and Whitaker, R. (1995). AKADAM: Eliciting user knowledge to support participatory ergonomics. *International Journal of Industrial Ergonomics, Special Issue: 'Participatory Ergonomics'*, **15**(5), 345–365.
- Miguez, S.A., Hallbeck, M.S. and Vink, P. (2012). Participatory ergonomics and new work: Reducing neck complaints in assembling. *Work*, **41**, 5108–5113.
- Moir, S. and Buchholz, B. (1996). Emerging participatory approaches to ergonomic interventions in the construction industry. *American Journal of Industrial Medicine*, **29**, 425–430.
- Montreuil, S. and Laville, A. (1986). Cooperation between ergonomists and workers in the study of posture in order to modify work conditions. In: E.N. Corlett, J.R. Wilson and I. Manenica, eds., *The Ergonomics of Working Postures*. London, U.K.: Taylor & Francis Group, pp. 293–304.
- Moore, J.S. (1994). Flywheel Truing—a case study of an ergonomic intervention. *American Industrial Hygiene Association Journal*, **55**, 236–144.
- Moore, J.S. and Garg, A. (1997). Participatory ergonomics in a red meat packing plant, part 1: Evidence of long-term effectiveness. *American Industrial Hygiene Association Journal*, **58**, 127–131.
- Moore, J.S. and Garg, A. (1998). The effectiveness of participatory ergonomics in the red meat packing industry: Evaluation of a corporation. *International Journal of Industrial Ergonomics*, **21**, 47–58.
- Mossink, J.C.M. (1990). Case history: Design of a packing workstation. *Ergonomics*, **33**, 399–406.
- Mumford, E. (1991). Participation in systems design – what can it offer? In: B. Shackel and S.J. Richardson, eds., *Human Factors for Informatics Usability*. Cambridge, U.K.: University Press, pp. 267–290.
- Nagamachi, M. (1995). Requisites and practices of participatory ergonomics. *International Journal of Industrial Ergonomics (Special Issue: 'Participatory Ergonomics')*, **15**(5), 371–379.
- Nagamachi, M. and Tanaka, T. (1995). Participatory ergonomics for reengineering in a chemical fibre company. In: *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*. Santa Monica, CA, pp. 766–770.
- Noro, K. (1991). Concepts, methods and people. In: K. Noro and A.S. Imada, eds., *Participatory Ergonomics*. London, U.K.: Taylor & Francis Group, pp. 3–30.
- O'Brien, D.D. (1981). Designing systems for new users. *Design Studies*, **2**, 139–150.
- Parimalam, P., Premalatha, M.R., Padmini, D.S. and Ganguli, A.K. (2012). Participatory ergonomics in redesigning a dyeing tub for fabric dyers. *Work*, **43**, 453–458.



- Riordan, C.M., Vandenberg, R.J. and Richardson, H.A. (2005). Employee involvement climate and organisational effectiveness. *Human Resource Management*, **44**, 471–488.
- Rivalis, I., van Eerd, D., Cullen, K., Cole, D.C., Irvin, E., Tyson, J. and Mahood, Q. (2008). Effectiveness of participatory ergonomic interventions on health outcomes: A systematic review. *Applied Ergonomics*, **39**, 342–358.
- Rosecrance, J.C. and Cook, T.M. (2000). The use of participatory action research and ergonomics in the prevention of work-related musculoskeletal disorders in the newspaper industry. *Applied Occupational and Environmental Hygiene*, **15**(3), 255–262.
- Ruohomaki, V. (1995). A simulation game for the development of administrative work processes. In: D. Saunders, ed., *The Simulation and Gaming Yearbook*. London, U.K.: Kogan Page, pp. 264–270.
- Seim, R. and Broberg, O. (2010). Participatory workspace design: A new approach for ergonomists? *International Journal of Industrial Ergonomics*, **40**, 25–33.
- Shaikh, I., Jayaram, U., Jayaram, S. and Palmer, C. (2004). Participatory ergonomics using VR integrated with analysis tools. In: R.G. Ingalls, M.D. Rossetti, J.S. Smith and B.A. Peters, eds., *Proceedings of the 2004 Winter Simulation Conference*, New York, pp. 1746–1754.
- Silverstein, B.A., Richards, S.E., Alcsér, K. and Schurman, S. (1991). Evaluation of in-plant ergonomics training. *International Journal of Industrial Ergonomics*, **8**, 179–193.
- Snow, M.P., Kies, J.K., Neale, D.C. and Williges, R.C. (1996). Participatory Design. *Ergonomics in Design*, **4**(2), 18–24.
- St-Vincent, M., Chicoline, D. and Beaugrand, S. (1998). Validation of a participatory ergonomics process in two plants in the electrical sector. *International Journal of Industrial Ergonomics*, **21**, 11–21.
- St-Vincent, M., Fernandez, J., Kuorinka, I., Chicoine, D. and Beaugrand, S. (1997). Assimilation and use of ergonomic knowledge by non-ergonomists to improve jobs in two electrical product assembly plants. *Human Factors and Ergonomics in Manufacturing*, **7**(4), 337–350.
- Sundin, A., Christmansson, M. and Larsson, M. (2004). A different perspective in participatory ergonomics in product development improves assembly work in the automotive industry. *International Journal of Industrial Ergonomics*, **33**, 1–14.
- Teiger, C. and Laville, A. (1987). How ergonomic methods can be used by non-ergonomists. *Proceedings of the Third International Occupational Ergonomics Symposium. Applied Methods in Ergonomics*, Zadar, Yugoslavia, pp. 14–16.
- Van Eerd, D., Cole, D., Irvin, E., Mahood, Q., Keown, K., Theberge, N., Village, J., St-Vincent, M. and Cullen, K. (2010). Process and implementation of participatory ergonomic interventions: A systematic review. *Ergonomics*, **53**(10), 1153–1166.
- Van der Schaaf, T.W. and Kragt, H. (1992). Redesigning a control room from an ergonomics point of view: A case study of user participation in a chemical plant. In: H. Kragt, ed., *Enhancing Industrial Performance*. London, U.K.: Taylor & Francis Group, pp. 165–178.
- Vink, P., Koningsveld, E.A.P. and Molenbroek, J.F. (2006). Positive outcomes of participatory ergonomics in terms of greater comfort and higher productivity. *Applied Ergonomics*, **37**, 537–546.
- Westlander, G., Viitasara, E., Johansson, A. and Shahnavaz, H. (1995). Evaluation of an ergonomics intervention programme in VDT workplaces. *Applied Ergonomics*, **26**(2), 83–92.
- Wilson, J.R. (1991a). Design Decision Groups – a participative process for developing workplaces. In: K. Noro and A. Imada, eds., *Participatory Ergonomics*. London, U.K.: Taylor & Francis Group, pp. 81–96.
- Wilson, J.R. (1991b). A framework and a foundation for ergonomics? *Journal of Occupational Psychology*, **64**, 67–80.
- Wilson, J.R. (1995a). Ergonomics and participation. In: J.R. Wilson and E.N. Corlett, eds., *Evaluation of Human Work: A Practical Ergonomics Methodology*. 2nd and revised edition. London, U.K.: Taylor & Francis Group, pp. 1071–1096.
- Wilson, J.R. (1995b). Solution ownership in participative work design: The case of a crane control room. *International Journal of Industrial Ergonomics*, **15**, 329–344.
- Wilson, J.R. (2005). Participatory ergonomics. In: J.R. Wilson and E.N. Corlett, eds., *Evaluation of Human Work*, 3rd edn. London, U.K.: Taylor & Francis Group, pp. 943.
- Wilson, J.R. and Grey, S.M. (1990). But, what are the issues in work design? In: C. Haslegrave, J.R. Wilson, E.N. Corlett and I. Manenica, eds., *Work Design in Practice*. London, U.K.: Taylor & Francis Group, pp. 7–15.
- Wilson, J.R. and Grey Taylor, S.M. (1995). Simultaneous engineering for self directed work teams implementation: A case study in the electronics industry. *International Journal of Industrial Ergonomics*, **16**(4), 353–365.

- Wilson, J.R. and Haines, H.M. (1997). Participatory ergonomics. In: G. Salvendy, ed., *Handbook of Human Factors and Ergonomics*, 2nd edn. Chichester, U.K.: John Wiley & Sons, pp. 490–513.
- Zalk, D.M. (2001). Grassroots ergonomics: Initiating an ergonomics program utilizing participatory techniques. *The Annals of Occupational Hygiene*, **45**(4), 283–289.
- Zink, K.J. (1996). Continuous improvement through employee participation. Some experiences from a long-term study in Germany. In: V.O. Brown Jr. and H.W. Hendrick, eds., *Human Factors in Organisational Design and Management*. London, U.K.: Elsevier Science, pp. 155–160.

---

# 35 Cost–Benefit Analysis for Human Factors

*Emile Tompa, Mike Foley and Dwayne Van Eerd*

## CONTENTS

|                                                                                 |     |
|---------------------------------------------------------------------------------|-----|
| Introduction.....                                                               | 907 |
| Negative Consequences of Poor Work System Design: The Burden of MSDs .....      | 909 |
| Scope of Burden Measurement .....                                               | 909 |
| Losses at the Individual Level.....                                             | 909 |
| Losses at the Family, Community and Societal Level .....                        | 910 |
| Role of Burden Studies .....                                                    | 911 |
| Evidence on the Burden of MSDs.....                                             | 912 |
| Methods and Issues in the Economic Evaluation of Ergonomics/Human Factors ..... | 913 |
| CBA and Monetary Measures of Health.....                                        | 914 |
| Human Capital .....                                                             | 914 |
| Friction Cost Approach .....                                                    | 915 |
| Revealed Preference Approach .....                                              | 915 |
| Stated Preference Approach.....                                                 | 916 |
| CEA, CUA and Non-Monetary Measures of Health.....                               | 917 |
| Selecting a Health Measure.....                                                 | 918 |
| Adjustments for Time Preference, Inflation and Uncertainty .....                | 918 |
| Evaluation Decision Rules and Summary Statistics .....                          | 919 |
| Perspective of an Analysis.....                                                 | 921 |
| Equity Issues .....                                                             | 921 |
| Evidence on the Financial Merits of Ergonomic Interventions .....               | 921 |
| Evidence Synthesis from a Systematic Review .....                               | 921 |
| Review of Recent Studies.....                                                   | 925 |
| Discussion and Summary.....                                                     | 925 |
| References.....                                                                 | 927 |

## INTRODUCTION

Organisations regularly face challenging financial decisions in their effort to get the most out of the funds they have available for their operations. This is as true in the realm of ergonomics/human factors investments as it is for other areas of organisational activities. At a broader, societal level, consideration of the resource implications of alternative programmes is equally as critical, as is the distribution of costs and consequences of programmes across the various stakeholders affected. Consequently, complete information on the resource implications (both costs and consequences) of alternatives from different stakeholder perspectives is critical to inform decision making in the realm of ergonomics/human factors. Essentially, economic evaluation is an integral part of programme evaluation, whether the programme under review is focussed on physical or mental health and well-being of workers. Some types of outcomes may pose more measurement challenges for economic analysis, particularly in areas where measurement tools to capture productivity impacts

are not well-developed (e.g. presenteeism) or where adverse consequences can be externalised to different stakeholders (e.g. health-care costs borne by the public sector). Also posing a measurement challenge is latent health conditions that may surface only in the long run (e.g. repetitive strain injuries, illnesses associated with chronic stress). Nonetheless, the principles of economic evaluation are still applicable and concerted efforts to adequately capture the key resource implications can provide invaluable information to the resource allocation decision-making process. Some might say that it is imperative to consider economic matters in the area of ergonomics/human factors – at the worker, employer and societal level – since it is not possible to invest in all programmes that are proven effective. In the short term, information on the resource implications of alternatives helps get the most out of expenditures by identifying the most cost-effective programmes. In the long run, it can help achieve a higher level of labour-market engagement of working-age adults, and ultimately enhance social welfare, by identifying the mix of ergonomics/human factors intervention programmes with the greatest value. Furthermore, through standardised economic evaluation information, decision makers at the organisational level can make comparisons across intervention types and benchmark their performance against appropriate peers.

Undertaking economic evaluations of ergonomics/human factors intervention programmes can be a challenge for a number of reasons: the policy arenas associated with ergonomics/human factors are complex, with multiple stakeholders and sometimes conflicting incentives and priorities; there are substantial differences in the perceptions of health risks associated with work experiences amongst the various stakeholders; there is sometimes a lack of consensus about what should be counted as a cost or benefit of intervening or not intervening; in some jurisdictions there are multiple providers of indemnity and medical care coverage, making it difficult to capture the full cost of workplace injuries and illnesses, and the benefits of their prevention; and lastly there are industry-specific human resource practices (e.g. hiring of temporary workers and farming out of services) that can make it difficult to identify the full extent of the injury and illness burden.

In this chapter, we focus on musculoskeletal disorders (MSDs) physical ergonomics and MSDs as a case study, but the principles of economic evaluation are broadly applicable to ergonomics/human factors. MSDs provide a good example of a consequence of inadequate work or system design that has individual and organisational implications. What sets MSDs apart from acute trauma injuries is their sometimes multifactorial nature and long latency, which can present unique challenges in evaluating the costs and consequences of alternative programmes to prevent them. Specifically, studies often have short measurement time frames, and as a result it is not uncommon to have insufficient power to detect a change attributable to an intervention over the measurement period. These challenges are further aggravated by new forms of work arrangements in which place of work is less clearly delineated, and hence attribution to work exposure is difficult to make (Table 35.1).

This chapter provides a starting point and some insights related to the aforementioned challenges, with reference to the case example of MSDs. The next section discusses methods of burden measurement and evidence on the magnitude of the MSD burden. It is followed by a section on

---

**TABLE 35.1**

**Key Terms**

*Resource implications of alternatives:* A term used to describe the sum of costs and consequences of programme options to mitigate or reduce a burden.

*Costs and consequences:* A phrase used to describe the valued inputs and outputs of a programme. The term *costs and benefits* is sometimes used, but this generally refers to values measured in a monetary metric, particularly in terms of health outputs of a programme.

*Economic evaluation:* A consideration of the costs and consequences of two or more programme options to mitigate or reduce a burden.

---

economic evaluation methods. The subsequent section highlights evidence on the economic merits of ergonomics programmes from the peer-reviewed literature. We end with a discussion and summary of the role of economics in programme evaluation and investment decision making, continuing with the focus on MSDs and ergonomics.

## NEGATIVE CONSEQUENCES OF POOR WORK SYSTEM DESIGN: THE BURDEN OF MSDs

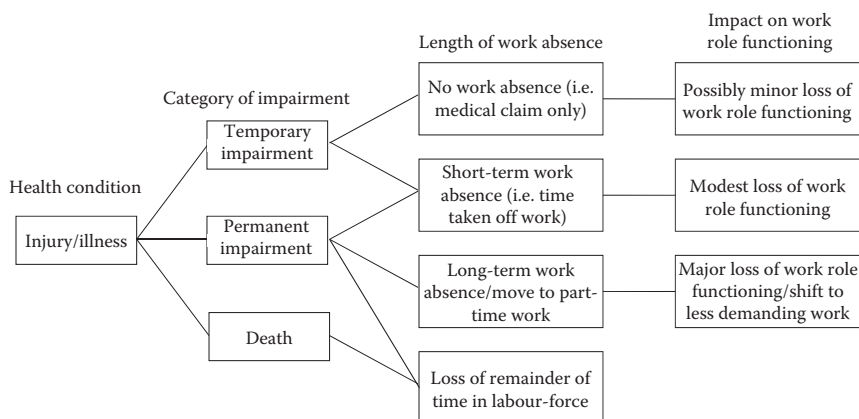
### SCOPE OF BURDEN MEASUREMENT

Often the awareness of the need to entertain a change in policy, programme or practice in workplace ergonomics arises through tracking the burden of work-related injuries/illnesses. Burden tracking may be informal, such as monitoring adverse health incidents and associated disability days at the employer level, or formalised in a societal-level burden of disease study. The latter is a study that measures the total loss of healthy time (i.e. morbidity and mortality) from a particular health condition, the costs of treating individuals with the condition and the impact of the condition in terms of undesirable consequences such as lost productivity. With formal burden studies, effort is made to capture all costs associated with morbidity and mortality from a particular injury/illness (or injuries/illnesses in general), including losses at the individual, family, community and societal levels.

### LOSSES AT THE INDIVIDUAL LEVEL

Figure 35.1 presents a schema that classifies the work impact of injury or illness experienced by a worker. The schema distinguishes between injury/illness (e.g. low back pain), impairment (e.g. lumbar mobility), lengths of work absence and impact on work role functioning (i.e. work disability). Work disability may result in income loss, but this may or may not be replaced by disability benefits. Factors that bear on such losses include the nature of the injury/illness, the availability of insurance coverage, whether a claim is filed and other factors such as the ability and willingness of the employer to accommodate the injured/ill worker.

Not depicted in the schema is the possibility of long-term labour-market earnings losses due to job loss or loss of promotion not directly attributable to work disability. These may occur for a variety of reasons such as the stigma associated with being an injured/ill worker, other worker–employer



**FIGURE 35.1** Schema for categorizing the work impact of injury or illness. (Adapted from Tompa, E. et al. 2008a.)

relationship issues and injured/ill worker loss of skills while off work recovering. Such losses may be associated with any category of impairment, work absence or work disability since they are not directly related to the severity of the injury or illness, but simple due to the fact of being an injured worker and filing a claim.

In the schema, impact on work role functioning is depicted through categories for heuristic purposes. The category of 'minor loss of role function' may include short-term impacts on presenteeism and absenteeism such as lower levels of engagement for 1 or 2 days due to pain and discomfort from low back pain. The category of 'modest loss of work role functioning' may include longer periods of loss that require formal temporary work accommodations or extended period of time off work. The category of 'major loss of work role functioning' connotes more serious, sometime permanent work disability that compromises a worker's ability to continue in the same type of employment.

The schema does not adequately depict the temporal aspect of adverse health conditions and ensuing impairment and disability. For example, for some injuries and illnesses there may be long latencies before they manifest. This is particularly true for some types of MSDs. Essentially, the biological effect of exposure may remain sub-acute for a long period of time. The injury/illness may manifest itself even after the removal/reduction of the hazard or exposure, unlike many acute injury events. There is also the greater probability of re-injury with MSDs not found with other injury types. In some cases, a health condition may only manifest as injury/illness after retirement or exit from the labour force, in which case they will not have an impact on work role functioning but rather on other social roles. Additionally, some health conditions may result in episodic or varying levels of impairment, and yet others may be degenerative. Consequently, impairment levels may have varying levels of severity in different time periods after first detection, and hence temporally varying impacts on work absences and work role functioning.

In addition to labour-market earnings losses, there can be other losses at the individual level not identified in the schema. For example, impairment from a work injury/illness may result in non-work disability (i.e. disability in other social roles such as home maintenance and parenting). There may also be out of pocket costs associated with health loss for services and products purchases to support recovery not reimbursable by insurance. Non-work losses may also include loss of health-related quality of life, sometimes described as the intrinsic value of health (see Bowling 1999 for an in-depth treatment of this topic). Examples include (1) chronic pain related to an injury or illness, (2) anger and depression related to job loss and impaired role functioning and (3) loss of enjoyment of leisure activities.

### **LOSSES AT THE FAMILY, COMMUNITY AND SOCIETAL LEVEL**

The individual-level losses described earlier – loss of income, loss of role functioning (e.g. work, home, community, leisure roles), out of pocket costs and loss of health-related quality of life – are not a full accounting of possible losses, since there can also be losses at the family, community and societal level. At the family and community level, the impact of individual health loss can affect family and community time use, including role substitution, reduced labour-market engagement and reduced community engagement. Family and community members may take time off work to care for the injured/ill worker or fulfil social roles normally upheld by the worker, such as home maintenance and parenting. The impact on family and community may be short- or long term. In the long run, adult outcomes may be different for children, such as curtailment of educational attainment due to financial and social constraints. Ensuing labour-market earnings may be affected accordingly. Family and community members' quality of life may also be affected in the short- and long run.

At the employer level, there may be output and productivity losses. In the short run, employer-level losses may be larger than the individual's productivity losses due to the nature of the production process (e.g. ability to replace the worker, team production, time sensitivity of output)



**TABLE 35.2**  
**Schema of Burden Categories across Stakeholders**

| Individual                                                                                                                                                                                                                                                                                                                            | Family and Community                                                                                                                                                                                                                                              | Employer                                                                                                                                                                                                                                           | Society                                                                                                                                                                                                        |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Labour-market earnings</li> <li>• Wage replacement benefits</li> <li>• Payroll benefits associated with labour-market earnings</li> <li>• Out-of-pocket expenses</li> <li>• Work role engagement</li> <li>• Engagement in other social roles</li> <li>• Intrinsic value of health</li> </ul> | <ul style="list-style-type: none"> <li>• Time use of family and community members</li> <li>• Quality of life of family and community members</li> <li>• Family income and savings</li> <li>• Adults outcomes of children</li> <li>• Community outcomes</li> </ul> | <ul style="list-style-type: none"> <li>• Productivity and output</li> <li>• Recruitment, training and replacement costs</li> <li>• Health insurance expenses</li> <li>• Disability insurance expenses</li> <li>• Other payroll expenses</li> </ul> | <ul style="list-style-type: none"> <li>• Productivity and output</li> <li>• Population health and quality of life</li> <li>• Capital accumulation, investment and related productivity implications</li> </ul> |

(see Nicholson et al. 2006 for a comprehensive discussion). Employer-level losses may be mitigated in the long run, even if the injured/ill worker is off work for an extended period of time, if the employer is able to recruit and train a replacement worker. Employer- and societal-level productivity losses in the short- and long run will depend on a number of factors; key variables are the size of the injury/illness burden across all workers in a particular time period, and the level of unemployment in the economy. At the employer, system and societal levels, there are also health-care costs, social support costs and programme administrative costs to include in the account of the burden.

Table 35.2 provides an overview of categories of values (by stakeholder) to consider for inclusion in burden of injury/illness studies. Customarily, only a subset of these categories is included due to the measurement challenges involved in quantifying them. The key ones generally included are healthcare/insurance expenses and labour-market earnings losses of the injured/ill worker (the latter is considered a proxy for employer productivity and output losses). The societal-level category is often thought of as the sum of values across stakeholders. A basic issue to address in aggregating values across stakeholders is the removal of any double counting of values. A more controversial issue is how to aggregate productivity and output losses at the societal level. Some economists hypothesise that such losses are only temporary, and that in the long run there are no losses. This is the premise underlying the friction-cost approach. A more traditional approach is to assume that losses at the aggregate level are enduring. This is the premise underlying the human capital approach. We describe these two approaches in more detail later in the chapter. An even more challenging issue, rarely even mentioned in burden studies, is the impact of large health burdens on economic general equilibrium and relative prices. This latter topic is one for only economic experts to grapple with! Whilst this is informed by MSDs in particular, the points in Table 35.2 are generalisable to other types of costs of failure to implement an ergonomics/human factors approach to work design.

## ROLE OF BURDEN STUDIES

Though burden studies do not measure the effectiveness of programmes for reducing the burden or their resource implications, these types of studies serve an important information role. They provide insights into the magnitude of the health and productivity losses and their costs to society. This information can be used to assess how the burden may have changed over time or how a burden from a particular health condition compares to other burdens (e.g. to measure progress in reducing burdens by monitoring trends over time). It can also help decision makers with priority setting. Burdens that are particularly onerous draw attention to the need to: (1) increase funding for programme options known to reduce the burden; (2) evaluate the

effectiveness and resource implications of programmes not yet evaluated and/or (3) invest in research to identify approaches to reducing the burden.

Burden estimates are typically reported for a specific calendar year and are based on costs in that year for all individuals diagnosed with or living with a particular condition. These aggregate costs are also referred to as prevalence costs because they encompass costs for individuals across the work disability trajectory, including new cases and the long-term disabled. Burden studies can also cost incidents longitudinally, starting from onset, and only include new cases. The time period for these longitudinal or incidence cost studies ranges from several months to the individual's lifetime. These two general types of burden studies are not directly comparable because of differences in the time periods measured and individuals included.

## EVIDENCE ON THE BURDEN OF MSDs

MSDs have received a lot of attention in the occupational health and safety (OHS) literature because they are the primary type of work-related injury and the key reason for filing a worker's compensation claim. Consequently, the burden of work-related MSDs is a concern. In one study, the MSD burden in the United States (for both work and non-work-related conditions) was identified as the leading cause of disability, with one in four Americans having a MSD condition requiring medical treatment (Jacobs et al. 2008). The burden amounted to \$849 billion in direct costs of healthcare and indirect costs of lost wages in 2004 (comprising a total of 7.7% of gross domestic product). For low back pain alone, the indirect costs in the United States were \$7.4 billion in that year (Dagenais et al. 2008).

Low back pain appears to be a large and growing burden in developed and developing countries around the world. In fact, it is the most prevalent of musculoskeletal conditions, affecting 75%–80% of all individuals at some point over their lifetime (Mody and Wolf 2003). Low back pain is most frequent in the working age population (Woolf and Pfleger 2003). In the United States, prevalence for the population as a whole is 15%–20% and in Europe it is 25%–45% (Mody and Wolf 2003).

A systematic review of low back pain burden studies by Dagenais et al. (2008) highlights the prevalence and economic burden of low back pain across several developed countries. The study reports on data from Australia, Belgium, Japan, Korea, the Netherlands, Sweden, the United Kingdom and the United States. Unfortunately, different methods were used to estimate direct and indirect costs across studies, so direct comparison is difficult. For example, of the 27 studies included in the review, 18 took a societal perspective, whereas 9 took a private insurer perspective. Nine considered direct costs only, nine indirect only and nine both. For indirect costs estimation, some used the human capital approach, others the friction cost approach and some did not specify the method. Indirect costs considered included sick leave, presenteeism, early retirement, household productivity loss and inactivity. The first three fall under the category of productivity and output losses, whereas the latter two under losses associated with non-work roles.

Even though there are differences in methods across studies, some overall patterns are worth noting. For example, the key direct healthcare costs, ranked by proportion of direct costs, were physical therapy (17%), inpatient services (17%), pharmacy (13%) and primary care (13%). Lost work output/productivity was the largest overall cost, in most cases amounting to over 80% of total costs (6 out of 8 studies that considered both direct and indirect costs). As a result of its high prevalence and large burden, much attention has been given to the prevention and treatment of low back pain, particularly in sectors where lifting is important part of some work tasks, such as construction, manufacturing, mining, forestry and healthcare.

Several studies have documented the burden of musculoskeletal injuries in monetary and non-monetary terms for different jurisdictions. The World Health Organization (WHO) published an extensive study on the global burden of MSDs (World Health Organization 2003), with the purpose of drawing attention to the importance of non-fatal health conditions and developing a common approach to measuring burdens, particularly for use in economic evaluations. The opening address

**TABLE 35.3****Key Terms**

*Burden of disease study:* A study that measures the total morbidity and mortality from a particular injury/illness or injuries/illnesses in general, or their impacts in terms of unfavourable consequences including the cost of treating the victims.

*Health-related quality of life:* Sometimes described as the intrinsic value of health, it refers to the effect of health on one's ability to live a meaningful and rewarding life. In a more formal measurement context, it refers to the construct underlying a set of measures used to quantify changes in health states associated with healthcare programmes.

*Direct costs:* In the context of burden studies, it usually refers to the costs closely attributable to a health condition such as medical and rehabilitation costs. In the context of an OHS programme, it usually refers to the internal cost of the programme in terms of the resources expended by the organisation initiating the programme.

*Indirect costs:* In the context of burden studies, it usually refers to the labour productivity effects attributable to a health condition and more distal effects such as on the family and community. Similarly, in the context of OHS programmes it refers labour productivity and more distal effects.

*Prevalence cost study:* A type of burden of disease study that considers the total morbidity and mortality, or their impacts in terms of unfavourable consequences, in a particular year regardless of time of onset.

*Incidence cost study:* A type of burden of disease study that considers the total morbidity and mortality, or their impacts in terms of unfavourable consequences, over the entire disease trajectory associated with new cases arising in a particular year.

*Presenteeism:* The effects of ill health on work engagement while at work. It is a counterpoint to absenteeism, which considers lost output and productivity losses due to absence from work.

of the WHO study notes that non-communicable diseases are the cause of 40% of death around the world. MSDs are also a major cause of morbidity, having important implications for health and quality of life. The European Agency for Safety at Work published a report on work-related MSDs in the European Union (Schneider and Irastorza 2010), with the aim of highlighting trends in MSDs, as well as the causes and circumstances associated with them. Lastly, the American Academy of Orthopaedic Surgeons published a report on the burden of MSDs in the United States (Jacobs et al. 2008), with the intent of providing broader insights into the burden and encouraging more research in the area (Table 35.3).

## METHODS AND ISSUES IN THE ECONOMIC EVALUATION OF ERGONOMICS/HUMAN FACTORS

Burden studies provide a first impression of the potential gains in health and related consequences to be realised if all cases of a particular injury/illness (or all injuries and illnesses) were eliminated. As such, they represent a maximum that could be gained. To evaluate the actual resource implications of known programmes requires an economic evaluation. An economic evaluation is a comprehensive evaluation of the costs and consequences of two or more alternative programmes. It is an approach to determining which alternative provides the greatest net value. The peer-reviewed OHS literature provides a broad range of programmes directed at reducing various types of MSDs that have been evaluated in terms of both their effectiveness and their resource implications. We draw on this literature in this section as we describe the methods and practice of economic evaluation of OHS programmes.

Resources (both costs and consequences) can be measured in monetary terms, but the money metric may not accurately capture the true value of health. Therefore, health is often measured with other metrics. In the health technology assessment field, it is customary to use non-monetary measures of health. In contrast, the OHS field tends to use monetary measures almost exclusively. Different types of economic evaluation methods have been developed to allow for non-monetary measures of health. Specifically, there are three different types of economic evaluation methods, with the metric used for health being the key difference between them. Cost–benefit analysis (CBA)

measures health in monetary terms, cost-effectiveness analysis (CEA) measures health in natural units (e.g. pain reduction, disability days averted) and cost-utility analysis (CUA) measures health in quality-adjusted health units in which level of morbidity and time in a particular health state are combined. All three approaches, CBA, CEA and CUA, measure costs in monetary terms. All three also use the monetary metric for non-health consequences. Next we review specific approaches to measuring the value of health in CBA, CEA and CUA.

## **CBA AND MONETARY MEASURES OF HEALTH**

Monetary measures of values are generally taken from market prices, but can also be identified through non-market means such as surveys. There are three common market approaches to measuring the value of health, namely the human capital approach, the friction cost approach and the revealed preference approach. The principal non-market approach relies on stated preferences about individuals' willingness to pay/accept captured through survey methods.

### **HUMAN CAPITAL**

The central premise of the human capital approach is an assumption that the value of health is primarily its human capital for use in productive activities (Drummond et al. 2005; Tompa et al. 2008a). The focus is often exclusively on output from paid labour-force activity. To estimate the value of lost output, absence time from an occupational role is multiplied by the value of time (its price weight) in that role. Implicit in this calculation is that the wage value of time off work due to poor health is a good measure of lost output at the employer and societal levels, that is the person is not replaced in their work role, and output loss endures until return to work. Factors such as the nature of the production process (e.g. team production, time sensitivity of output) and replaceability of the injured worker are generally not factored into the calculations. In some instances, the value of time in non-work roles is considered, particularly for home production. The intrinsic value of health is not considered.

For price weights of occupational time, different rates may be used. In the evaluation of workplace interventions, actual salaries are often used, whereas for population studies average salaries are used. Price weights should include the full compensation package (i.e. wages plus benefits). The present value of earnings losses until return-to-work or retirement is estimated using a discount rate if an absence spans longer than a year. For long-duration absences, rather than assuming that wages remain constant for the injured/ill worker, one can adjust earnings to reflect the standard lifetime earnings trajectory. For example, if a young adult earns minimum wage and becomes permanently and fully disabled, one could assume some earnings growth over the career that would be based on a counterfactual of what the person would have been earning if they had not been injured.

For individuals not in the paid labour force – youth, students, homemakers, retirees – it is not clear how best to value their time in poor health. One possibility for non-paid occupational roles (e.g. home maintenance) is to estimate what it would cost to pay someone to do that task (replacement cost) or what the person would be paid if they were in the paid labour market (see Drummond et al. 2005 for details).

There are four key concerns regarding the human capital approach. First, wage rates may not accurately reflect the marginal product of worker time due to market imperfections. Second, its focus on occupational output as the only value of health is too narrow by many accounts. Third, in its simplest form (where actual wage values are used) the approach places greater value on the time of individuals with greater earnings, and lesser value on the time of individuals with lesser earnings. Fourth, a strong assumption commonly made when using this metric is that societal output losses due to long-term disability are enduring. In reality, if a worker is absent for a long time, the person would likely be replaced with a worker who would eventually be equally as productive. Long-run output and productivity at the employer level may be negligible. This might also be the case at the

---

**TABLE 35.4**  
**Case Example 1**

Bultmann et al. (2009) evaluated a coordinated and tailored work rehabilitation programme compared to conventional case management through a randomised controlled trial undertaken in Denmark. The key outcome was return-to-work. Participants were workers on sick leave for 4–12 weeks due to an MSD. A CBA from the societal perspective considered intervention costs, productivity loss based on the human capital approach and health-care costs. The intervention was found to significantly reduce the number of sickness absence hours, and was cost beneficial. The net present value at 12 months was US\$10,666 per person.

---

societal level. Though there would still be a loss of potential output, since a return to full employment would be at a lower output level due to the loss of the worker's productive capacity.

Because the human capital approach takes a very narrow view of the value of health, it is not commonly used as the sole measure of health in programme evaluation. The exception is the OHS field where intervention studies often take an employer's perspective whose concerns are often focused on maintaining productivity and output (Tompas et al. 2009a). Weil (2001) suggests using the human capital approach to measure the value of healthy time in the paid labour force, and using a non-monetary measure to capture the value of healthy time in non-work roles and the intrinsic value of health (Table 35.4).

### **FRICTION COST APPROACH**

The friction cost approach was developed to address the assumption underlying the human capital approach that societal output losses due to long-term disability are enduring. With the friction cost approach, it is contended that societal output losses exist only in the short run, even if some individuals with a health condition are off work for extended periods of time. This is the case because individuals amongst the ranks of the unemployed eventually are recruited to replace those who become injured or ill. The period during which losses are incurred is known as the friction period. The friction cost approach might be viewed as an aggregation approach rather than a distinct way of estimating the productivity and output losses due to injury and illness. Essentially, estimating productivity and output losses at the individual level can be undertaken with any particular methodological approach and then aggregated based on assumptions underlying the friction cost approach. For example, during the friction period the human capital approach may be used to estimate productivity and output losses.

There are some limitations with the friction cost approach. It assumes that the number of people employed is essentially independent of the labour supply and that there is much less than full employment at all times. Excess unemployment is required for losses to be mitigated in the long run by providing a sufficiently large number of individuals available to take up the position made vacant by those who become injured or ill. Even if this is the case, the friction period may vary over the business cycle and over time as the unemployment rate varies. Furthermore, even if unemployment is somewhat above the full employment level at all times, labour supply may still exert some influence on employment bargains and the total number hired. Given these factors, it is not clear what value to use as the friction period in a burden study. If studies use different friction periods for interventions executed in different time periods, comparability between studies can become a challenge. Possibly one can undertake sensitivity analysis with a range of values.

### **REVEALED PREFERENCE APPROACH**

The revealed preference approach is based on an evaluation of the choices people make in the market (e.g. through the choice of a job with known health risks in exchange for higher pay). It is

premised on the notion that individual command a compensating wage differential to take on health risks. Additional assumptions include full information of health risks for job seekers, and no power imbalances between them and employers. If this is the case, statistical methods can be used to identify the value of health risks using data on jobs that have different levels of risk and the wages paid for those jobs. Much of the revealed preference literature on the compensation wage differentials has focussed on the risk of fatality. The resulting calculations are for the statistical value of a human life. Some work has also been undertaken on the risk of morbidities. Several extensive reviews provide good insight into the extent of this literature (Dorman 1996; Dorman and Hagstrom 1998; Viscusi 1993).

The revealed preference approach is not often used in economic evaluations for several reasons. First, developing measures for various health conditions requires identifying revealed preference situations with particular health risks and then collecting and analysing data from them. Values for a full complement of health risks would be difficult to assemble due to the absence of data and opportunities to collect them. As noted, most studies to date have focused on the risk of death in an occupational context rather than the risk of morbidity, whereas morbidity is an important aspect of MSDs. A second shortcoming is that health risk values found in different studies have been, in some cases, dramatically inconsistent. Third, it is difficult to know the full range of features of different jobs that bear on the wage differentials identified. They may be due to undesirable features other than health risk, such as the griminess of a particular occupation. Lastly, there may be factors present that bias the health risk values identified through revealed preferences, such as lack of information on the part of job seekers about the health risks of different occupations, and power imbalances between job seekers and employers, particularly for recent immigrants who may have language barriers or lack knowledge about legislation, rights and the going wage level for specific occupations in a particular country.

### STATED PREFERENCE APPROACH

The stated preference approach relies on direct elicitation of values through survey methods (Drummond et al. 2005; Tompa et al. 2008a). This method is also called contingent/stated valuation because individuals are asked directly about the values they ascribe to alternatives. The approach is commonly used in environmental assessments, and has also been used in the health technology assessment field. Questions in the survey ask respondents to state the amount of money they are willing to pay to receive a health benefit, or the amount they are willing to accept to live with a health condition. Willingness to pay (WTP) identifies the maximum amount an individual would be willing to pay for a health improvement. Willingness to accept (WTA) identifies the minimum amount an individual would be willing to accept to bear a burden. In programme evaluation, if the sum of WTP of all affected individuals exceeds the costs of implementing an alternative, then the intervention is deemed to be worth undertaking. Applied to the context of ergonomics, WTP questionnaires can be used to value programmes to reduce MSDs in monetary terms.

There are several shortcomings with WTP/WTA measures. The key shortcoming is that they are sensitive to the ability to pay. Thus programmes benefiting those with more disposable income may be given priority over programmes benefiting those of more modest means. Other concerns include whose preferences to elicit, and how broadly or narrowly to cast the questions about the value of consequences. A broad question would inquire about the willingness to pay for all consequences, including consideration of labour-market earnings associated with adverse health conditions. A more restricted WTP approach might focus only on the intrinsic value of health, which would then require capturing productivity/output consequences separately through other means (e.g. by using market prices). Another shortcoming particularly relevant to MSDs is that it may be difficult for healthy respondents to imagine the consequences of living/working with pain if the latency period before onset is long. It may also be difficult to compare studies from different jurisdictions if concepts of



what is considered ‘normal’ levels of pain while working are culturally specific. Lastly, there are asymmetries between WTP and WTA due to valuation being affected by the starting endowment.

**CEA, CUA AND NON-MONETARY MEASURES OF HEALTH**

The term CEA is used to refer to economic evaluations that measure health in natural units. These can include intermediate and final outcome measures such as clinical measures, particular symptoms or general health status (e.g. blood pressure, pain, discomfort, cases of low back pain, self-reported health status). Measures can be specific to a particular health condition or a measure of general health. They may be intermediate proxies for downstream health outcomes (e.g. pain, discomfort) or direct measures of end-state health. The choice of measure to use depends on the purpose and context of a study (e.g. efforts to reduce low back pain or carpal tunnel syndrome).

One of the key concerns with CEA is the limitations on comparability. Only studies using similar outcome measures can be compared. Even when studies use apparently similar measures, they may not be fully comparable due to the use of different measurement protocols, for example pain being measured using different questionnaires with different scales. General health measures such as the SF-12 are more broadly applicable and comparable and have been tested for construct validity and reliability. However, such measures may be less responsive to health changes from an intervention than purpose-specific measures, particularly in the short run. Another concern is that non-health consequences need to be captured through other, preferably monetary measures, if they are to be included in an analysis. A third concern is the need for an external yardstick to assess the monetary value of a unit of health outcome. Essentially, the decision maker will at some point be confronted with the need for information on how much an employer or society is willing to pay for a unit of health (Table 35.5).

The term CUA refers to evaluations using health-related quality of life measures that combine quality (i.e. level of morbidity) and quantity (i.e. length of time in a health state). Such measures are meant to more fully capture the complexity of health than natural units. Specific types of CUA measures include quality-adjusted life-years (QALYs) and variants such as healthy year equivalents (HYEs), EuroQol and disability-adjusted life-years (DALYs) (Drummond et al. 2005; Gold et al. 1996; World Health Organization 2011). There are also health classification systems known as preference-based multi-attribute health status classifications systems, such as quality of well-being, health utility index (HUI) and EQ-5D which can be used in conjunction with data on the length of time in health state to construct health-related quality of life measures.

Several issues arise with the construction of health-related quality of life measures. A key issue is how to weight health-related quality of life measures when aggregating units within and across individuals. The convention is to weight units equally, regardless of their distribution across individuals, though this may not necessarily reflect societal preferences (e.g. a unit of health may be of higher societal value if received by a child than an elderly person). A third issue is that health-related

**TABLE 35.5**  
**Case Example 2**

Meijer et al. (2006) undertook a randomised control trial in the Netherlands of a multidisciplinary treatment for sick-listed workers with upper extremity MSDs. The intervention consisted of psychological and physical sessions with a medical specialist, a psychologist, a physiotherapist and an occupational therapist. The key intervention outcomes included severity of complaints and return-to-work. Over a 12-month period, the intervention was found to be effective in improving physical functioning and reducing severity of complaints, but was similar to usual care in terms of return-to-work. The CEA, undertaken from the societal perspective, was not cost-effective compared to usual care based on statistical significance. In terms of actual values, the incremental cost per extra half-day return-to-work required an investment of €39 over 12 months, and the incremental cost per extra unit decrease in pain €66.

**TABLE 35.6****Case Example 3**

Squires et al. (2011) assessed the cost-utility of three return-to-work interventions for workers with MSDs. The interventions were (1) a workplace intervention, (2) a physical activity and education intervention and (3) a physical activity, education and workplace visit intervention. A Markov model was used with data from various sources. The CUA found all three interventions had a cost per QALY below £3000 compared to usual care within the United Kingdom. Based on this information, the interventions are likely to be considered worthwhile to the UK National Health Service due to the low cost per QALY.

quality of life measures only capture the value of health to the patients benefiting from a programme. Not captured are benefits to others such as family and community. In economics, these are termed health externalities. They are associated with contagious diseases and sentiments such as altruism and parentalism (e.g. the value one places on good health for others). A fourth issue the concern about how and where worker time costs are captured (i.e. labour-market earnings associated with different health states) and aggregate productivity consequences associated with health outcomes. The consensus seems to be to measure them separately in monetary terms, but that risks double counting because respondents to preference elicitation exercises are unlikely to be able to meaningfully separate out the intrinsic value of health from its value associated with labour-market activity (Table 35.6).

In the earlier example, the Markov model, which is like a probability tree, is used to delineate the multiple future health states that can follow from a particular level of health risk exposure and their related probabilities, based on the health risk exposure in each alternative. As noted, the approach requires data from various sources on future health states and probabilities since the intervention is not evaluated in a particular setting. The value of this approach is that one can stylise the magnitude of expected outcomes without undertaking an on-site evaluation, which can be time consuming and costly undertaking. The downside of the approach is that the stylised outcomes identified may not be generalisable.

### SELECTING A HEALTH MEASURE

Choosing between one of the three types of economic evaluations (CBA, CEA and CUA) should be based on the objective of the intervention and the question being addressed by the study. These in turn are influenced by the nature of the key outcome variable and the relevant perspective(s) to be considered. For example, if a key perspective is that of a private sector firm, CBA might be preferred if one of the objectives of an intervention is to reduce insurance costs through improved OHS performance. Including more than one type of evaluation is not uncommon since each can provide different insights into the merits of an intervention. For example, one could undertake a CEA to better capture the health outcomes that are not readily translatable into a monetary metric (e.g. pain reduction, disability days averted), and a CBA in which health outcomes are proxied through some monetary measure (e.g. human capital approach).

### ADJUSTMENTS FOR TIME PREFERENCE, INFLATION AND UNCERTAINTY

Both costs and consequences, whether measured in monetary units or other metrics, must be adjusted for the timing of their flow using an appropriate interest rate. This is known as discounting and is critical if resource flows occur over more than 1 year. Discounting is based on the premise that the timing of resource flows has implications for their value (known as time preference). In economic theory, it is assumed that individual preferences are such that more immediate access to resources is given a higher value than deferred access.

If prices are expressed in nominal values from different calendar years, one should also adjust for inflation to make them comparable in terms of purchasing power. It is generally advisable to first make adjustments for inflation and then discount the costs and consequences expressed in constant dollars (or another monetary metric) to a base year.

Many jurisdictions stipulate the discount rate at which public sector investments are to be discounted. For the private sector, firms may have their own specific rate used for project investments. The discount rate recommended by several researchers is 3% (Gold et al. 1996; Goossens et al. 1999; Haddix et al. 2003), but because a number of studies in the past used 5%, it is recommended to consider both rates in an analysis. Undertaking a sensitivity analysis using a range of rates from 0% to 10% is also recommended (Haddix et al. 2003). This is particularly important in cases where costs are realised early and positive consequences, such as health improvement, are realised only well into the future. In such cases, a high discount rate may result in a low present value of the positive consequences, and are thus more likely to be outweighed by the more immediate costs of an intervention.

Varying the discount rate to test the robustness of results is important, but generally not sufficient to address all aspects of uncertainty within a study. In general, there are many other aspects of a study that are less than certain, and therefore need to be evaluated in terms of their robustness to different assumption. Such robustness testing is known as sensitivity analysis. Sensitivity analyses should also present the reasons for the different values tested. There are various approaches to conducting sensitivity analysis across multiple dimensions (e.g. multi-way sensitivity analysis, threshold analysis, scenario analysis, probabilistic sensitivity analysis) (see Claxton et al. 2005; Drummond and Sculpher 2005).

Projections into the future are often an area of uncertainty and warrant sensitivity analysis. For example, in an evaluation of an ergonomics programme, the costs of averted MSDs are often based on their lifetime health and productivity implications. But these costs are often identified through projections into the future; consequently, it would be appropriate to test different values. Future projections of costs and consequences associated with ongoing interventions are another key area of uncertainty.

## EVALUATION DECISION RULES AND SUMMARY STATISTICS

Summary measures for CBA, CEA and CUA are generally reported as a ratio of the cost per monetary benefit, natural unit or QALY. The latter is a metric commonly used in the health-care sector. As noted in Table 35.7, CUA health measures combine time and quality of a particular health state. The measure is broadly applicable to different health interventions but may not be sensitive to subtle changes in health states.

Whether undertaking CBA, CEA or CUA, the ratio of cost and consequences is to be calculated as an incremental/marginal cost (relative to a comparator such as the standard programme) per incremental/marginal benefit, natural unit or QALY (again relative to a comparator). Calculating this ratio can be a challenge, particularly with CEA and CUA. Good guidance on decision rules is provided by Drummond et al. (2005) and Hoch and Dewa (2008). Table 35.8 provides a summary. Because both numerator and denominator in CBA are in monetary units, what changes in values are placed in one versus the other can be inconsistent across studies making ratios across studies difficult to compare. For example, reductions in workers' compensation costs may be perceived as a cost in one study and a benefit in another. If workers' compensation costs are placed in the numerator of a CBA ratio, the ratio will be different than if they are placed in the denominator. Alternative summary measures used in CBA less prone to this issue are net present value (NPV) and payback period. These measures also have issues. NPV is affected by the scale of an intervention, for example an intervention in a small firm will have a small NPV compared to the same intervention scaled up for a large firm, *ceteris paribus*. This issue may be addressed by scaling NPV by the relevant unit in the analysis such as a claim, case or worker (e.g. NPV per worker). Payback period does not provide insight into the costs and consequences that follow after the intervention costs are recovered.

**TABLE 35.7**  
**Health Measures Used in Economic Evaluations**

| Type of Evaluation | Health Measure                                                 | Definition                                                                                                                                                | Limitations                                                                                                                                                                                                                                                                                                                                                |
|--------------------|----------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CBA                | Human capital approach                                         | Monetary value of health estimated by the wage value of time off work                                                                                     | Focus on output and productivity as the primary value of health is limited<br>Wage rates may not accurately capture the marginal product of a worker<br>Assumption that societal productivity losses from long-term absences are enduring may not be realistic                                                                                             |
|                    | Friction cost approach                                         | Aggregation approach that considers productivity and output losses for only a friction period                                                             | Assumption that there is always excess unemployment may not always be the case<br>Comparability across studies can be a problem if different friction periods are used                                                                                                                                                                                     |
|                    | Revealed preference approach (compensating wage differentials) | Monetary value of health states identified through the trade off people make between bearing health risks in exchange for compensating wage differentials | Not always clear what consequences are captured in the statistical values identified<br>Statistical values for all health states are not available and computationally demanding to calculate<br>Values found in different studies are inconsistent<br>Labour-market imperfections may distort values                                                      |
|                    | State preference approach (willingness to pay/accept)          | Monetary value of health states identified by asking individuals to state their preferences in terms of their willingness to pay/accept                   | Values are sensitive to ability to pay/accept so results may be biased towards health conditions affecting higher income earners<br>Difficult to determine how broadly to cast the question about preferences and what to capture through other means                                                                                                      |
| CEA                | Natural units                                                  | Health states measured in natural units that reflect intermediate and final health outcomes                                                               | Comparisons limited to studies using similar measures<br>Generic measures may be less responsive to the interventions than purpose-specific measures<br>Non-health outcomes need to be captured through other means<br>Health externalities are not captured<br>Monetary value of a unit of health outcome needs to be identified to make policy decisions |

(Continued)

**TABLE 35.7 (Continued)**  
**Health Measures Used in Economic Evaluations**

| Type of Evaluation | Health Measure                 | Definition                                                       | Limitations                                                                                                                                                                                                                                                                                                                  |
|--------------------|--------------------------------|------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CUA                | Health-related quality of life | Health states measured in time units adjusted for health quality | Quality and quantity assumed separable and divisible<br>Measures may not be sensitive to subtle changes in health states<br>Non-health outcomes need to be captured through other means<br>Health externalities not considered<br>Monetary value of a unit of health outcome needs to be identified to make policy decisions |

**PERSPECTIVE OF AN ANALYSIS**

Most economic evaluations of workplace interventions found in peer-reviewed journals are conducted from the perspective of the firm or company. A focus on the company perspective may be warranted if the firm is the key decision maker, but omitting consideration of the costs and consequences experienced by other stakeholders may overlook critical costs and consequences. There is a strong normative argument for considering a broad, societal perspective and for considering the distribution of costs and consequences across various stakeholders. Specifically, the fact that there are multiple stakeholders affected by ergonomics/human factors issues (firms, workers and their families, unions, healthcare providers, insurers, society) suggests that costs and consequences borne by all the stakeholders ought to be included in the analysis. This is the norm in other economic evaluation contexts where there are multiple stakeholders (e.g. environmental impact assessment). A broad perspective does not preclude providing information on other perspectives. In fact, a disaggregation of costs and consequences would be invaluable as it would provide insight into their distribution.

**EQUITY ISSUES**

Economic evaluation is focussed on allocative efficiency. Inherent in the methodology is a need to compare and aggregate costs and consequences across individuals and across different stakeholder groups. As a result, there are equity implications of interventions that ought to be explicitly considered in an evaluation. Equity issues are commonly placed under two broad categories – distributive equity and procedural equity (Table 35.9). The former refers to the fairness of the allocation of costs and consequences, whereas the latter refers to the fairness of the decision or allocation process. All equity constructs have inherent values embodied in them, so none can be assessed exclusively through scientific principles. Within the broad constructs of distributive and procedural equity there are many rival notions that have been proposed. A summary of the key ones found in the literature is presented by Culyer and Tompa (2008).

**EVIDENCE ON THE FINANCIAL MERITS OF ERGONOMIC INTERVENTIONS**

**EVIDENCE SYNTHESIS FROM A SYSTEMATIC REVIEW**

Though the literature on ergonomic and other MSD injury prevention interventions is quite extensive, the economic implication of such interventions is considered in fewer studies. A systematic literature review on the topic (Tompa et al. 2010) identified 35 peer-reviewed ergonomic intervention

**TABLE 35.8**  
**Types of Economic Evaluations and Related Decision Rules**

| Type of Economic Evaluation | Type of Economic Comparison Being Evaluation Made                   | Standard Decision Rule                                                                                                                                                                        | Summary Measure(s)                                        | Example                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Issues                                                                                                                                                                                                                                                                                                                                                            |
|-----------------------------|---------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CBA                         | CBA incremental cost (DC) compared to incremental benefit (DB)      | If $DC < DB$ then the intervention is worth undertaking                                                                                                                                       | Cost-benefit ratio<br>Net present value<br>Payback period | Lahiri et al. (2005): For an office ergonomics programme consisting of lumbar pads, backrests and a back school workshop the net savings per year were \$70,441 with savings per worker of \$111. The benefit-to-cost ratio was 84.9 and the payback period was 0.5 months (2002 US dollars)                                                                                                                                                                        | Difficult to determine what to put in the numerator versus the denominator of a CBA ratio<br>Net present value may be affected by the scale of an intervention<br>Payback period does not provide insight into the costs and consequences that follow after the intervention costs are recovered<br>Studies using different natural units are not easily compared |
| CEA                         | CEA incremental cost (DC) compared to incremental natural unit (DE) | If the value a unit of effect is worth the cost as identified by the cost-effectiveness ratio, then the intervention is worth undertaking, but only if there is money available in the budget | Cost-effectiveness ratio                                  | Loisel et al. (2002): For a disability management intervention consisting of a clinical intervention combined with occupational intervention (Sherbrooke model), at mean 6.4 years follow-up, the relative cost per days on full benefit (DFB) (compared to the standard care arm) was (−\$67.50) per DFB saved for the clinical arm, (−\$88.40) per DFB saved for the occupational arm and (−\$63.50) per DFB saved for the Sherbrooke arm (1991 Canadian dollars) | Incremental costs and/or incremental effects may be negative relative to the comparator, making it difficult to interpret the finding of an evaluation                                                                                                                                                                                                            |
| CUA                         | CUA incremental cost (DC) compared to incremental natural unit (DU) | If the value of a QALY is worth the cost as identified by the cost-utility ratio, then the intervention is worth undertaking, but only if there is money available in the budget              | Cost-utility ratio                                        | Kermode et al. (2003): For a Q fever vaccination programme, increasing vaccination uptake from 65% to 100% amongst meat industry workers resulted in a cost per QALY of \$6294; increasing vaccine uptake from 0% to 20% amongst agricultural industry workers resulted in a cost per QALY of \$7984 (2001 Australian dollars)                                                                                                                                      | It is not always clear what is captured in a QALY because there are different ways to estimate a QALY<br>There is a need to identify a maximum dollar value for a QALY to use in the decision rule<br>Incremental costs and/or incremental QALYs may be negative relative to the comparator, making it difficult to interpret the finding of an evaluation        |



**TABLE 35.9****Key Terms**

*Human capital approach:* An approach to valuing healthy time in terms of its human use in productive activities. It is often translated into a monetary metric by multiplying the time off work due to poor health by the wage rate for that occupation.

*Friction cost approach:* An approach to aggregating the value of healthy time across individuals. The key assumption underlying the approach is that productivity and output losses are only experienced at the firm and societal level for limited period of time, known as the friction period.

*Revealed preference approach:* An approach to identifying values through market behaviours. They are generally identified through estimates of compensating wage differentials paid to workers for bearing health risks.

*Stated preferences:* An approach to identify the value of health by asking individuals to identify the maximum amount they are willing to pay for health improvement or the minimum amount they would be willing to accept to live with a health condition.

*Health-related quality of life measures:* Measures that combines quality (i.e. level of morbidity) and quantity (i.e. length of time in a health state) of health. Quality-adjust life-years (QALYs) is one of the most widely known measure of this type.

*Health externalities:* Health values not captured in a measure of the value of health. The term generally refers to the value of health to those other than the recipient of the health benefits. Three aspects of health values are often identified as external to many measures – values to others due to altruism, paternalism and reductions in contagion.

*Discounting:* An adjustment made to resource flows due to their differential timing. It is based on the premise that the timing of resource flows has implications for their value, known as time preference.

*Sensitivity analysis:* A testing of the robustness of results of a study to different value of one or more key parameters that are not certain. This is undertaken to ensure results are not dependent on assumptions made about the value of these parameters.

*Distributive equity:* In general relates to ethical judgements about the fairness of the distribution of values such as income, wealth, health, costs and benefits.

*Procedural equity:* In general relates to ethical judgements about the fairness of the process by which decisions or allocation are reached.

studies with economic evaluations, though only 16 were retained in the synthesis based on their quality (i.e. only those receiving a medium or high-quality score in the quality assessment phase studies were retained).<sup>\*</sup> In the discussion section, we provide a detailed description of the key methodological shortcomings identified in this literature. The 16 remaining studies were in 6 industry sectors: administrative and support services, health care, information and culture, manufacturing and warehousing, retail and trade and transportation. We summarise the finds for four sectors next. For all other sectors in which studies were identified, there was insufficient evidence on the cost-effectiveness of ergonomic interventions.

Most studies (10 of 16) were undertaken in the United States. Two studies were in Canada, two in Australia and one each in Sweden and the Netherlands. Four studies were participatory ergonomic interventions and three were ergonomic education programmes on back health and safety. They covered a wide range of features and levels of intensity, some of which were labour-intensive (e.g. participatory ergonomic teams) and others which were capital intensive (e.g. mechanical patient lifts in hospitals, highly adjustable office chairs and workstation modifications). The predominant economic outcomes considered in evaluations were workers' compensation expenses, including both the wage replacement and health-care components of these expenses, and the monetary value of absenteeism. In terms of perspective, all adopted a firm perspective.

<sup>\*</sup> In the study, quality was assessed by two reviewers using a 14-item quality assessment tool that had questions related to four broad categories: (1) overarching issues that frame the purpose of the study and the nature of the intervention; (2) study design and issues related to evaluating the intervention's effectiveness; (3) measurement and analytic issues related to the economic analysis and (4) issues related to the discussion and interpretation of results.

In the administrative and support sector, the synthesis identified moderate evidence that ergonomic interventions are worth undertaking on the basis of their financial merits. Three medium and one high-quality studies were identified (Derango et al. 2003; Lahiri et al. 2005; Rempel et al. 2006). These interventions targeted work station equipment and training for office workers. Equipment included highly adjustable chairs, lumbar pads and backrests and track ball and arm-boards with computer use. Training included appropriate use of equipment and back school workshops. Two studies had more than one intervention arm including a control (both used regression modelling techniques to control for confounders), while a third study was a before–after study without a separate control. The three studies included in this sector all undertook a cost–benefit analysis, and considered insurance and productivity consequences.

For the health-care sector, there was also moderate evidence that ergonomic interventions are worth undertaking for economic reasons, based on three medium-quality studies (Chhokar et al. 2005; Collins et al. 2004; Evanoff et al. 1999). Interventions included the introduction of mechanical patient lifts in two cases and the implementation of a participatory ergonomic team in the other. The target populations were individuals working in a hospital setting, such as nurses, nurses' aides and orderlies. Study designs were before–after without controls, two of which used regression modelling techniques to control for confounders. Regarding the economic evaluation method employed, two studies undertook a cost–benefit analysis while the other conducted a cost–consequence analysis (i.e. costs and consequences are analysed separately rather than jointly). Only insurance consequences were considered in the economic analyses.

Studies in the manufacturing and warehousing sector provided strong evidence that ergonomic interventions are worth undertaking for their financial merits. There were two medium-quality evaluations and three high-quality ones (Abrahamsson 2000; Halpern and Dawson 1997; Lahiri et al. 2005; Lanoie and Tavenas 1997),\* and all concluded that the ergonomic interventions were cost-effective in this sector. The interventions were focused on a broad range of MSD injury prevention measures for individuals working with machinery. In three cases, the interventions were participatory, while in the other two instances they consisted of engineering controls and workstation modifications. All were before–after uncontrolled studies, with one using regression modelling techniques to control for confounders. Four studies undertook a cost–benefit analysis, and one was a partial analysis that only considered insurance consequences.

Studies in the transportation sector provided limited evidence that such interventions result in economic returns, based on two medium- and one high-quality evaluation (Daltroy et al. 1997; Tuchin and Pollard 1998; Versloot et al. 1992). The two medium-quality studies found the interventions to be cost-effective, whereas the high-quality one did not. The three studies focused on back injury prevention. All were randomised controlled trials though were not blinded. Regression modelling and analysis of variance were undertaken to assess the difference between and within groups. With regards to the economic evaluation component, each study undertook a different type of analysis. One study was a partial analysis, the second a cost-consequence analysis and a third a cost–benefit analysis. Insurance and productivity consequences were considered.

Across all sectors, there were a total of 6 high-quality studies and 10 medium-quality ones. Of the 16 studies, all but one found the interventions to be worth undertaking based on their financial merits. Consequently, when considering evidence across all sectors, there is strong evidence that ergonomic interventions result in economic returns for the firm.

Only four studies were participatory ergonomic interventions. Three of them were medium quality and one was of high quality (Abrahamsson 2000; Evanoff et al. 1999; Halpern and Dawson 1997; Lanoie and Tavenas 1997). This results in moderate evidence that participatory ergonomic interventions are worth undertaking based on their financial returns for the firm.

---

\* Lahiri et al. (2005) contained two high quality evaluations.

## REVIEW OF RECENT STUDIES

We review several recent economic evaluation studies undertaken in several industries – administrative and support services, health care and manufacturing and warehousing – to provide a glimpse of recent studies in the field. One study evaluates a participatory ergonomics programme to prevent low back and neck pain with computer users at four Dutch companies (Driessen et al. 2012). The programme consisted of working group meetings to brainstorm about, evaluate and prioritise risk factors of low back and neck pain. The top three priorities had to be implemented by the group within a 3-month period. A clustered randomised controlled trial with 12-month follow-up served as the basis for effectiveness and economic evaluation. The focus was on health care and productivity costs. The programme was found to be not cost-effective or cost-beneficial from the societal perspective. From a firm perspective, the monetary loss was €78 per person.

Two studies evaluate efforts to reduce MSDs associated with patient handling amongst care workers. Nelson et al. (2006) evaluate a multifaceted ergonomics programme implemented in seven facilities in south eastern United States. The intervention included six programme elements, with the key ones being state-of-the-art equipment, peer support and a no lift policy. A before–after design with no control is used to evaluate effectiveness and economic merits over a 9 month before and 9 month after time period. The focus was on workers' compensation and productivity costs. The intervention was found to be effective and cost beneficial, with a payback period of 3.75 years. The second patient handling study evaluated a ceiling lift programme implemented at three facilities in British Columbia Canada (Alamgir et al. 2008). A before–after analysis with no concurrent controls includes a time period of 4 years before and 6 years after implementation. The focus was on workers' compensation and lost productivity costs (disability days). The study found the intervention to be effective in reducing MSDs due to patient handling. The payback period was estimated to be between 2.06 and 3.20 years. Both studies take the firm or care facility perspective.

The last study evaluates a participatory ergonomics programme introduced at a car parts manufacturer in Ontario, Canada (Tomba et al. 2009b). The study uses a before–after design (18 months before and 23 months after) with statistical control for contextual factors through regression modelling techniques. Here too the focus is workers' compensation and lost productivity costs (disability days). Both cost-effectiveness and cost–benefit analysis are undertaken. In terms of the former, the ratio was found to be \$12.06 per disability day. The benefit-to-cost ratio was 10.6, with a net present value of \$244,416, suggesting the intervention was worth undertaking from the firm perspective.

## DISCUSSION AND SUMMARY

Two key issues are apparent in the ergonomics literature: (1) few ergonomic intervention studies undertake an economic evaluation and (2) the intervention studies that do undertake economic analyses present a diversity of methodological approaches and quality with a large number of low-quality studies. Other reviews of the literature have come to similar conclusions (DeRango and Franzini 2003; Goossens et al. 1999; Niven 2002). The large number of low-quality studies is likely due to the focus in the OHS literature on effectiveness rather than cost-effectiveness. Also, undertaking economic evaluations of workplace interventions can be a challenge for a number of reasons, and there is little guidance available on how it should be done. Most methods are designed for use in a clinical, health-care setting, but a number of factors in the workplace setting are different than those in the clinical setting. We noted these challenges in the introduction and elaborate on them here: (1) the policy arena is complex, with multiple stakeholders and sometimes conflicting incentives and priorities; (2) there are substantial differences in the perceptions of health risks associated with work experiences amongst workplace parties, policymakers and other stakeholders; (3) there is a consequential lack of consensus amongst stakeholders about what, in principle, ought to count as a benefit or cost of intervening or not intervening; (4) the burden of costs and consequences may be borne by different stakeholders in the system; (5) there are multiple providers of indemnity and

medical care coverage in some jurisdictions, such that no one administrative data source accurately captures the full cost of work-related injury and illness, nor conversely, the benefits of their prevention; (6) industry-specific human resources practices (e.g. hiring temporary workers and self-employed contractors, outsourcing non-core activities) can make it difficult to identify all work-related injuries and illnesses and (7) there is a dearth of data available from organisations making it both challenging and expensive to obtain good measures. Even in relatively large organisations, the focus of their internal accounting systems often does not lend itself readily to identifying the full costs and benefits of an intervention. Specifically, the relevant data are not collected at the correct level of disaggregation and impacts on related production processes can be difficult to estimate *a priori*. Therefore, some impacts on the ecology of a plant, facility or the organisation as a whole may have to be left in a qualitative form.

Many studies in the literature rely on poor study designs for effectiveness evaluation, and these results also feed into the economic evaluation. The merits of different epidemiological designs are well documented (Lilienfeld and Stolley 1994; Rothman and Greenland 1998). Most workplace intervention researchers would argue that randomised controlled trials are the design of choice, but non-randomised controlled studies, interrupted time series and simple before–after studies can also be well executed and provide valuable information. Many OHS studies with economic analyses are based on before–after study design without randomisation or concurrent controls. The main concern with before–after studies is the attribution of all changes in the outcome of interest to the intervention, without accounting for confounding factors. In all study designs, but particularly with before–after designs, known or conjectured confounders that are not controlled for through randomisation ought to be controlled for in some other way, for example through multivariate modelling techniques using interrupted time series data. For example, a study by Tompa et al. (2013) uses aggregate weekly time series data at the plant level to estimate the impact of a participatory ergonomics intervention on a range of health and safety outcomes. A dummy variable in the regression models identifies the time period as before or after the introduction of the intervention, and the models control for the number of regular and overtime production hours in the week, number of new staff hires in the week and number of staff separations (terminations/resignations, layoffs) in the week.

A number of standard computational practices are overlooked in some analyses. For example, when the costs and/or consequences of an intervention are realised over more than a year, one should adjust for inflation and time preference. Data on inflation rates are readily available from most national statistical agencies. To adjust for time preference, discounting is required for both costs and consequences, even if consequences are not measured in dollars. Many jurisdictions stipulate the discount rate at which public sector investments are to be discounted. For the private sector, firms may have their own specific rate used for project investments.

Many studies consider only a limited subset of intervention costs and consequences. Most focus on work absence costs (primarily wage costs or workers' compensation wage replacement costs) and medical care costs. One concern with wage costs or workers' compensation wage replacement costs as the sole or primary outcome measure is that they do not capture the full set of costs and consequences, even from a firm's perspective. A range of indirect costs may be incurred by a firm that results in costs substantially larger than the direct absence costs. Furthermore, productivity may be affected even while an injured/ill worker is at work. To accurately assess productivity one needs to consider the nature of the production process and the product/service being produced since factors such as team production, time sensitivity and substitutability of a worker will affect output (Pauly et al. 2002). With regards to workers' compensation costs, they are a very poor measure of absence costs because they are simply transfers and do not capture the true cost of disability days. Additionally, workers' compensation claims do not reflect the full extent of work-related injuries and illnesses. Compensable injuries and illnesses may go unreported (Shannon and Lowe 2002), and others are not compensable. With accepted claims, not all losses may be fully compensated, for example costs such as a long-term lower earnings and loss of household productivity. Other costs may be externalised onto social support programmes such

as social security and publicly funded health care. Some absences may be reported under other compensation programmes such as a private disability insurance.

Most studies undertake cost–benefit analysis, and focus on output and productivity changes due to the intervention. Although financial outcomes and productivity issues may be of primary interest to most firms, the value of health to other stakeholders, particularly injured workers and their families includes much more. Missing in most financial measures of health are the intrinsic value of good health to workers and the value of health associated with the ability to better perform in other social roles.

The perspective taken matters for the workplace measures of health used in an evaluation. In fact, the perspective bears on all the costs and consequences considered in an economic evaluation. Most published studies take the firm perspective, considering only those costs and consequences experienced by the firm. But there is a strong case to be made for considering other perspectives, particularly those of the worker and system or society, as well as for a disaggregation of the costs and consequences by stakeholder in order to better understand their composition and distribution.

Another concern is that studies employ different approaches to the computation and analysis of costs and consequences, making it difficult to compare results across studies. For example, some studies with cost and consequences in monetary terms use net present value, others the payback period, yet others a cost–benefit ratio. A standard approach to computations following reference case such as Gold et al. (1996) or Tompa et al. (2008b) would greatly facilitate comparability. Variation in labour-market legislation and disability compensation programmes across jurisdictions may complicate comparability, but standardisation of analysis and reporting can still facilitate understanding of transferability and generalisability. Though the guidance provided by these sources tend to focus on physical health conditions, the basic principles of economic evaluation are generic and the approaches recommended can be extended to other types of ergonomics/human factors interventions that focus on other outcomes such as cognitive tasks. The OHS domain offers much knowledge and guidance, particularly around quantifying effects. The literature on presenteeism also provides insights into quantifying health-related engagement, performance and productivity.

In summary, information on economic burdens and the resource implications of alternative ergonomics/human factors interventions is important for policy decision making at the local and national levels, and is likely to matter even more in the future as populations in many developed countries continue to age, changes in health patterns continue to unfold, cost of work disability continues to increase and globalisation makes competitiveness even more intense. There are many challenges to undertaking economic evaluations, and these are magnified with new technologies and changing forms of work and work contracting. MSDs pose a particular problem in this context because they are sometimes multifactorial in nature and of long latency. Furthermore, economic information is useful only if one can discern the nature and quality of the evidence and the transferability/generalisability of the findings reported in studies. In this chapter, we reviewed methods and measures for burden and economic evaluation, and presented references for other sources of guidance that can serve to build a foundation for understanding and interpreting such studies. The focus on ergonomics/human factors provided a grounding and examples for the discussion of economic issues, as well as insights into the current state of the evidence and areas warranting further research.

## REFERENCES

- Abrahamsson, L. 2000. Production economics analysis of investment initiated to improve working environment. *Applied Ergonomics* 31(1):1–7.
- Alamgir, H., Yu, S., Fast, C., Hennessy, S., Kidd, C., Yassi, A. 2008. Efficiency of overhead ceiling lifts in reducing musculoskeletal injury among carers working in long-term care institutions. *Injury: International Journal of the Care of the Injured* 39:570–577.
- Bowling, A. 1999. Health-related quality of life: A discussion of the concept, its use and measurement background: The ‘quality of life’. Presented to the Adapting to Change Core Course, September 1999. <http://info.worldbank.org/etools/docs/library/48475/m2s5bowling.pdf> (accessed: 24 August 2012).



- Bultmann, U., Sherson, D., Olsen, J., Hansen, C. L., Lund, T., Kilsgaard, J. 2009. Coordinated and tailored work rehabilitation: A randomized controlled trial with economic evaluation undertaken with workers on sick leave due to musculoskeletal disorders. *Journal of Occupational Rehabilitation* 19:81–93.
- Chhokar, R., Engst, C., Miller, A., Robinson, D., Tate, R. B., Yassi, A. 2005. The three-year economic benefits of a ceiling lift intervention aimed to reduce healthcare worker injuries. *Applied Ergonomics* 36(2):223–229.
- Claxton, K., Sculpher, M., McCabe, C. et al. 2005. Probabilistic sensitivity analysis for NICE technology assessment: Not an optional extra. *Health Economics* 14:339–347.
- Collins, J. W., Wolf, L., Bell, J., Evanoff, B. 2004. An evaluation of a ‘best practices’ musculoskeletal injury prevention program in nursing homes. *Injury Prevention* 10(4):206–211.
- Culyer, A. J., Tompa, E. 2008. Chapter 13: Equity. In *Economic Evaluation of Interventions for Occupational Health and Safety: Developing Good Practice*, eds. E. Tompa, R. Dolinschi, A. Culyer, pp. 215–235. Oxford, U.K.: Oxford University Press.
- Dagenais, S., Caro, J., Haldeman S. 2008. A systematic review of low back pain cost of illness studies in the United States and internationally. *The Spine Journal* 8(1):8–20.
- Daltroy, L. H., Iversen, M. D., Larson, M. G. et al. 1997. A controlled trial of an educational program to prevent low back injuries. *The New England Journal of Medicine* 337(5):322–328.
- DeRango, K., Amick III, B. C., Robertson, M. M., Rooney, T., Moore, A., Bazzani, L. 2003. The productivity consequences of two ergonomic interventions. Working Paper #222. Toronto, Ontario, Canada: Institute for Work & Health.
- DeRango, K., Franzini, L. 2003. Economic evaluations of workplace health interventions: Theory and literature review. In *Handbook of Occupational Psychology*, eds. J. C. Quick, L. E. Tetrick, pp. 417–430. Washington, DC: American Psychological Association.
- Dorman, P. 1996. *Markets and Mortality: Economics, Dangerous Work, and the Value of Human Life*. New York: Cambridge University Press.
- Dorman, P., Hagstrom, P. 1998. Wage compensation for dangerous work revisited. *Industrial and Labour Relations Review* 52:116–133.
- Driessen, M., Bosmans, J., Proper, K., Anema, J., Bongers, P., van der Beek, A. 2012. The economic evaluation of a participatory ergonomics programme to prevent low back and neck pain. *Work: A Journal of Prevention, Assessment and Rehabilitation* 41:2315–2320.
- Drummond, M., Sculpher, M. 2005. Common methodological flaws in economic evaluations. *Medical Care* 43:5–14.
- Drummond, M., Sculpher, M. J., Torrance, G. W., O’Brien, B. J., Stoddart, G. L. 2005. *Methods for the Economic Evaluation of Health Care Programmes*, 3rd edn. Oxford, U.K.: Oxford University Press.
- Evanoff, B. A., Bohr, P. C., Wolf, L. D. 1999. Effects of a participatory ergonomics team among hospital orderlies. *American Journal of Industrial Medicine* 35(4):358–365.
- Gold, M. R., Siegel, J. E., Russell, L. B., Weinstein, M. C. 1996. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.
- Goossens, M. E., Evers, S. M., Vlaeyen, J. W., Rutten-an Mølken, M. P., van der Linden, S. M. J. P. 1999. Principles of economic evaluation for interventions of chronic musculoskeletal pain. *European Journal of Pain* 3:343–353.
- Haddix, A. C., Teutsch, S. M., Corso, P. 2003. *Prevention Effectiveness: A Guide to Decision Analysis and Economic Evaluation*. New York: Oxford University Press.
- Halpern, C. A., Dawson, K. P. 1997. Design and implementation of a participatory ergonomics program for machine sewing tasks. *International Journal of Industrial Economics* 20(6):429–440.
- Hoch, J., Dewa, C. 2008. Chapter 9: Kind of analysis and decision rule. In *Economic Evaluation of Interventions for Occupational Health and Safety: Developing Good Practice*, eds. E. Tompa, R. Dolinschi, A. Culyer, pp. 147–164. Oxford, U.K.: Oxford University Press.
- Jacobs, J. J., Andersson, G. B. J., Bell, J. E., Weinstein, S. L., Dormans, J. P., Gnat, S. M., Lane, N., Puzas, J. E., St. Clair, E. W., Yelin, E. H. 2008. The burden of musculoskeletal diseases in the United States: Prevalence, societal and economic costs: Executive summary. Rosemount, MN: American Academy of Orthopaedic Surgeons, p. 6. [www.usbjd.org](http://www.usbjd.org) (accessed 21 August 2012).
- Lahiri, S., Gold, J., Levenstein, C. 2005. Estimation of net-costs for prevention of occupational low back pain: Three case studies from the US. *American Journal of Industrial Medicine* 48:530–541.
- Lanoie, P., Tavenas, S. 1997. Costs and benefits of preventing workplace accidents: The case of participatory ergonomics. *Safety Science* 24(3):181–196.
- Lilienfeld, D. E., Stolley, P. D. 1994. *Foundations of Epidemiology*. New York: Oxford University Press.



- Loisel, P., Lemaire, J., Poitras, S., Durand, M. J., et al. 2002. Cost-benefit and cost-effectiveness analysis of a disability prevention model for back pain management: a six year follow up study. *Occupational and Environmental Medicine* 59(12):807–815.
- Meijer, E. M., Sluiter, J. K., Heyma, A., Sadiraj, K., Frings-Dresen, M. H. W. 2006. Cost-effectiveness of multidisciplinary treatment in sick-listed patients with upper extremity musculoskeletal disorders: A randomized, controlled trial with one-year follow-up. *International Archives of Occupational and Environmental Health* 79:654–664.
- Michelle, K., Yong, K., Hurley, S., Marmion, B. 2003. An economic evaluation of increased uptake in Q fever vaccination among meat and agricultural industry workers following implementation of the National Q Fever Management Program. *Aust NZ J Public Health* 27(4): 390–398. <http://dx.doi.org/10.1111/j.1467-842X.2003.tb00415.x>
- Mody, G. M., Woolf, A. D. 2003. A report on the global burden musculoskeletal disorders. Business briefing: European Pharmacotherapy Association, pp. 136–141. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.200.7777&rep=rep1&type=pdf>
- Nelson, A., Matz, M., Chen, F., Siddharthan, K., Lloyd, J., Fragala, G. 2006. Development and evaluation of a multifaceted ergonomics program to prevent injuries associated with patient handling tasks. *International Journal of Nursing Studies* 43:717–733.
- Nicholson, S., Pauly, M. V., Polsky, D., Sharda, C., Szrek, H., Berger, M. L. 2006. Measuring the effects of work loss on productivity with team production. *Health Economics* 15:111–123.
- Niven, K. J. 2002. A review of the application of health economics to health and safety in healthcare. *Health Policy* 61(3):291–304.
- Pauly, M. V., Nicholson, S., Xu, J. et al. 2002. A general model of the impact of absenteeism on employees. *Health Economics* 11:221–231.
- Rempel, D. M., Krause, N., Goldberg, R., Benner, D., Hudes, M., Goldner, G. U. 2006. A randomised controlled trial evaluating the effects of two workstation interventions on upper body pain and incident musculoskeletal disorders among computer operators. *Occupational and Environmental Medicine* 63(5):300–306.
- Rothman, K. J., Greenland, S. 1998. *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven.
- Schneider, E., Irastorza, X. 2010. OSH in figures: Work-related musculoskeletal disorders in the EU – facts and figures. European Agency for Safety and Health at Work, 184pp. <https://osha.europa.eu/en/publications/reports/TERO09009ENC>
- Shannon, H. S., Lowe, G. S. 2002. How many injured workers do not file claims for workers' compensation benefits? *American Journal of Industrial Medicine* 42:467–473.
- Squires, H., Rick, J., Carroll, C., Hillage, J. 2011. Cost-effectiveness of interventions to return employees to work following long-term sickness absence due to musculoskeletal disorders. *Journal of Public Health* 34(1):115–124.
- Tompa, E., Culyer, A. J., Dolinschi, R. 2008b. Chapter 14: Suggestions for a reference case. In *Economic Evaluation of Interventions for Occupational Health and Safety: Developing Good Practice*, eds. E. Tompa, A. J. Culyer, R. Dolinschi, pp. 235–244. Oxford, U.K.: Oxford University Press.
- Tompa, E., Dolinschi, R., de Oliveira, C. 2008a. Chapter 11: Consequences. In *Economic Evaluation of Interventions for Occupational Health and Safety: Developing Good Practice*, eds. E. Tompa, A. J. Culyer, R. Dolinschi, pp. 179–200. Oxford, U.K.: Oxford University Press.
- Tompa, E., Dolinschi, R., de Oliveira, C., Amick III, B. C., Irvin, E. 2010. A systematic review of workplace ergonomic interventions with economic analyses. *Journal of Occupational Rehabilitation* 20(2):220–234.
- Tompa, E., Dolinschi, R., de Oliveira, C., Irvin, E. 2009a. A systematic review of occupational health and safety interventions with economic analyses. *Journal of Occupational and Environmental Medicine* 51(9):1004–1023.
- Tompa, E., Dolinschi, R., Laing, A. 2009b. An economic evaluation of a participatory ergonomics process in an auto parts manufacturer. *Journal of Safety Research* 40:41–47.
- Tompa, E., Dolinschi, R., Natale, J. 2013. Economic evaluation of a participatory ergonomics intervention in a textile plant. *Applied Ergonomics* 44(3):480–487.
- Tuchin, P., Pollard, H. 1998. The cost-effectiveness of spinal care education as a preventive strategy for spinal injury. *Journal of Occupational Health and Safety – Australia New Zealand* 14(1):43–51.
- Versloot, J. M., Rozeman, A., van Son, A. M., van Akkerveken, P. F. 1992. The cost-effectiveness of a back school program in industry: A longitudinal controlled field study. *Spine* 17(1):22–27.
- Viscusi, W. K. 1993. The value of risks to life and health. *Journal of Economic Literature* 31:1912–1946.
- Weil, D. 2001. Valuing the economic consequences of work injury and illness: A comparison of methods and findings. *American Journal of Industrial Medicine* 40:418–437.

- Woolf, A. D., Pfleger, B. 2003. Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization* 81(9):649–656.
- World Health Organization. 2003. The burden of musculoskeletal conditions at the start of the new millennium. WHO Technical Report Series 919, 228pp. [http://whqlibdoc.who.int/trs/WHO\\_TRS\\_919.pdf](http://whqlibdoc.who.int/trs/WHO_TRS_919.pdf) (accessed: 24 August 2012).
- World Health Organization. 2011. Disability weights, discounting and age weighting of DALYs. [http://www.who.int/healthinfo/global\\_burden\\_disease/daly\\_disability\\_weight/en/index.html](http://www.who.int/healthinfo/global_burden_disease/daly_disability_weight/en/index.html) (accessed: 12 May 2012).

---

# 36 Integrating Ergonomics and Quality Concepts

*Jörgen A.E. Eklund and Paul H.P. Yeow*

## CONTENTS

|                                                                                              |     |
|----------------------------------------------------------------------------------------------|-----|
| Introduction.....                                                                            | 931 |
| Ergonomics and Quality Definitions and Historical Trend of Integration .....                 | 932 |
| Method .....                                                                                 | 934 |
| Framework for Integrating Ergonomics with Quality Concepts.....                              | 934 |
| Validation of Framework (Content Validity).....                                              | 934 |
| TQM versus Socio-Technical Systems Approach.....                                             | 936 |
| Customer Focus versus Ergonomics Stakeholder Model.....                                      | 937 |
| Quality Policy Deployment versus Participatory Ergonomics.....                               | 937 |
| ISO 9000 Quality Management System versus Psychosocial and Organisational Factors.....       | 938 |
| Process Orientation versus Systems Ergonomics.....                                           | 939 |
| Zero Defects versus Systems Ergonomics .....                                                 | 940 |
| Quality Value Chain versus Ergonomics Value Chains .....                                     | 940 |
| Continuous Improvement versus Participatory Ergonomics .....                                 | 940 |
| Quality Tools versus Ergonomics Principles and Tools .....                                   | 942 |
| Noise.....                                                                                   | 943 |
| Vibrations .....                                                                             | 943 |
| Climate .....                                                                                | 943 |
| Lighting .....                                                                               | 944 |
| Chemical Health Hazards.....                                                                 | 944 |
| Physical Workload.....                                                                       | 944 |
| Psychosocial and Organisational Factors .....                                                | 945 |
| Guidelines to Use the Framework.....                                                         | 945 |
| Case Studies.....                                                                            | 947 |
| Case Study 1: Continuous Improvements versus Participatory Ergonomics .....                  | 947 |
| Case Study 2: ISO 9000 versus Psychosocial and Organisational Factors .....                  | 948 |
| Comparison with Other Ergonomics and Quality Integration Methods, Frameworks and Models..... | 949 |
| Limitations of the Framework.....                                                            | 950 |
| Conclusion .....                                                                             | 951 |
| References.....                                                                              | 951 |

## INTRODUCTION

This chapter aims to present an overview of the relationships between ergonomics and quality, and to develop a framework to integrate ergonomics and human factors (E/HF) concepts with quality concepts (such as total quality management [TQM] and quality management systems). The aim is also to show how the approaches used in the two knowledge fields can mutually contribute to one another. Unlike the quality discipline, which is normally directly linked to the quality

department/function in organisations (Elg et al., 2011), E/HF does not have a direct link to any organisational department/function, and also lacks methods to be integrated into a company's value chain (Dul and Neumann, 2009; Dul et al., 2012). This poses a problem to ergonomists in integrating E/HF in their organisations. Thus, this chapter presents a framework for ergonomists to apply their knowledge in their organisations, particularly in collaboration with the quality function. There are several reasons for focusing on ergonomics and quality integration:

1. Ergonomics and quality disciplines have many similarities: both are multidisciplinary; have a basis in human reactions, preferences and needs; are relatively young and are largely of an applied nature (Eklund, 1997). Indeed, several research groups have developed research at the borderline between them (see Drury, 1997; Eklund, 1997; Carayon et al., 1999; Zink, 1999) and even proposed integrating the two (Eklund, 1997; Axelsson, 2000; Dul and Neumann, 2009).
2. People in general want to perform a good job with high-quality work results. To be able to produce good quality work is an important prerequisite for job satisfaction, learning and self-esteem and also for recognition from others (Deming, 1986). However, for this, they need preconditions and support that allow for good quality output, and these are addressed in the E/HF domain.
3. Many prior studies have identified quality deficiencies as being caused by poor ergonomics as the quality of human performance is affected by E/HF issues such as noise, vibration, temperature, lighting, chemicals, physical workload, postures, psychosocial work conditions and the work organisation (Eklund, 1999; Axelsson, 2000; Yeow and Sen, 2003, 2006; Erdinc and Yeow, 2011). However, many quality practitioners are still unaware of that ergonomics offers a new approach to improve quality, which provides ergonomists an opportunity to contribute by using the principles/methods presented in this book (*Evaluation of Human Work*) to address E/HF as well as quality issues.
4. Today, there is a strong trend towards measurements and evaluations of projects and different change initiatives. In particular, there is a strong emphasis on measuring economic consequences from not only quality improvements but also ergonomics and work environment improvements although these are sometimes not considered core activities for an organisation (Dul and Neumann, 2009; see Chapter 35 of this book).

This chapter summarises a large number of studies that confirm the strong relationship between ergonomics and quality performance, and it also shows benefits from integrating the two disciplines.

## **ERGONOMICS AND QUALITY DEFINITIONS AND HISTORICAL TREND OF INTEGRATION**

According to *The International Ergonomics Association's* definition 'Ergonomics (or human factors) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimise human well-being and overall system performance. Practitioners of ergonomics and ergonomists contribute to the planning, design and evaluation of tasks, jobs, products, environments and systems in order to make them compatible with the needs, abilities and limitations of people' (IEA, 2014). Wilson (2014a) emphasises that ergonomics is a systems approach that focuses on the interactions among people, tasks, interfaces, workplaces, environments and organisations at work (rather than on the individual components) to optimise economics, social and environmental outcomes. Based on these definitions, E/HF includes an individual perspective relating to the interaction between the individual and the technology, environment and organisation with safety, health and well-being as aims. This includes job satisfaction, pleasure, learning and personal development. In addition, there is a systems perspective,

relating to the human–systems interactions with organisational performance as the aim (productivity, efficiency, effectiveness, quality, innovativeness, flexibility, systems safety and security, reliability and sustainability) (Eklund and Bergman, 2003; see Chapter 1 of this book). Further, other societal outcomes may be included here (environment protection and sustainable development) (Hendrick, 2008; Dul et al., 2012; Wilson, 2014a).

Quality was earlier defined as ‘Fitness for use’ (Juran, 1974). Later, the definition was widened, that is ‘The quality of a product or service is its ability to satisfy the needs and expectations of the customers’ (Bergman and Klefsjö, 2003). It is often pointed out that it is desirable not only to satisfy but also to exceed the expectations of the customer, that is to surprise the customer positively. In the context of E/HF, the term customers can be used for consumers, who are referred to as external customers. The terms can also mean internal customers who are the manufacturing personnel in an industrial context. Naturally, all customers have expectations and needs regarding their interactions with products, which also include E/HF aspects of the products they use or in terms of the manufacturing process. These expectations and needs relate to the definitions of both ergonomics and quality, thereby confirming the overlap of the definitions.

There is a trend towards the integration of quality and ergonomics concepts. The quality concept has undergone a transition from (1) inspection to (2) quality control to (3) quality assurance to (4) continuous improvement (CI) and to (5) TQM (Golomski, 1994). In addition to TQM, another line of development is the ISO 9000 standard. The old paradigm of quality stated that improved quality could only be achieved at the cost of productivity losses (Deming, 1986), which is consistent with the view that quality is achieved through more intensive inspection. Further support for the old view was given by Fitts’ law: there is a trade-off effect between speed and accuracy of work, which has been interpreted as a trade-off between quality and productivity (Helander, 2006). However, results from companies, which adopted the newer view of improving quality by improving the manufacturing processes instead of focusing on quality inspection, show that the old paradigm is not valid. There is instead a positive relationship between good quality performance and productivity (Womack et al., 1990). This newer view has, of course, meant much for the quality strategies developed and applied, and for the attractiveness of quality issues in companies and organisations. Thus, the quality field is considered to be a strategic management issue, while E/HF is normally not. However, quality literature increasingly emphasises the role of humans in the transition to total quality (Deming, 1986; Lammermeyr, 1990; Dahlgaard, 2002). The reason is that the quality field has gained from experience that human aspects and human resource management principles are important preconditions for high-quality performance (Redman and Mathews, 1998; Tari and Sabater, 2006).

The field of E/HF has undergone a substantial change in scope and interest. Earlier studies were mainly involved in surveying the consequences of work design for humans and performance. Traditional inherent conflicts between management and unions earlier demanded that ergonomists develop ‘objective’ methods and threshold criteria for evaluation of existing work situations. There has always been an interest in an applied approach, developing ergonomics designs and solutions to ergonomics problems. Later, the issues of processes for ergonomics change and the role of ergonomists in these have been highlighted, along with an increased emphasis on E/HF interventions (Westgaard and Winkel, 1997). We can now identify a shift towards dealing with processes and interventions, as well as widening the type of issues at stake (Dul et al., 2012). Earlier focus on the work context has broadened to include leisure activities. Participatory ergonomics (PE) has developed, and an increasing interaction is taking place with the field of quality (Imada, 1991; Drury, 1997; Kogi, 2006, 2008; Caroly et al., 2010) and organisational design or macro-ergonomics (Hendrick and Kleiner, 2001). The latest development extends ergonomics broader and beyond the organisational perspective, that is to the overall systems perspective, which looks not only into the interaction of humans, technology, organisations and environments but also at all the stakeholders together with their goals (Wilson, 2014a,b). Quality and workers’ well-being are just parts of the overall system goals, with other goals such as productivity, profit and environment outcomes.

## METHOD

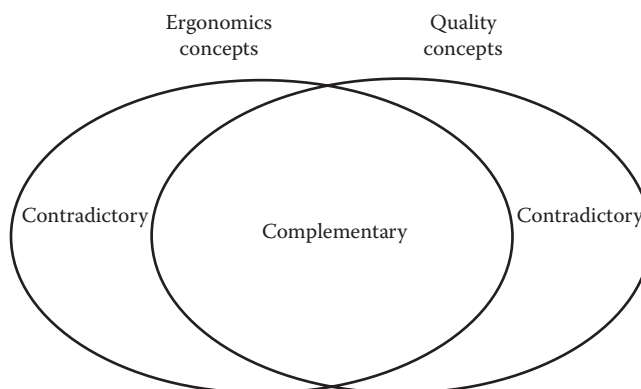
A framework was developed and validated based on literature review of 40 years of prior research on ergonomics and quality integration, which highlights opportunities for E/HF to be integrated with existing quality concepts. This method of validation, termed ‘content validity’, is often practised in social science and business studies. The review has its starting point in an industrial context but will expand to some extent into the service sector. Searches were made using tools including Science Direct, Ergonomics Abstracts, Emerald Library, Scopus, and Taylor & Francis Group Online, and focused on the quality and ergonomics terms used in the framework. The terms ‘ergonomics’ and ‘human factors’ were used interchangeably in the search (and in this paper), since they have the same meaning. The literature search included peer-reviewed journal articles, conference proceedings, books, theses, reports and website content. The literature included does not aim to give a complete list but instead a partial list of the main studies so as to provide the practitioner with an applicable knowledge of the integration. The literature gathered from the quality and ergonomics concepts was paired based on complementary and/or contradictory relationships. A complementary relationship is when the concepts go well together or help to better each other, whereas in contradictory relationships, the concepts are opposites or in disagreement with each other. Based on this, the pairs were classified into two areas (see Figure 36.1), that is a complementary area where there are opportunities for easy integration or already existing integration due to the complementary relationship between the ergonomics and quality concepts, and a contradictory area where there was no existing integration or integration was difficult due to the contradictory relationship between the two concepts. Comparisons were conducted between ergonomics and quality concept pairings, that is whenever there was a quality concept, there was a comparison with an ergonomics concept. Some paired concepts appeared in both areas due to different aspects/interpretations/views of the quality concept compared with the ergonomics concept.

## FRAMEWORK FOR INTEGRATING ERGONOMICS WITH QUALITY CONCEPTS

Table 36.1 shows the framework for integrating ergonomics with quality concepts. The paired quality and ergonomics concepts are presented in rows, which can be complementary or contradictory relationships or both. Explanations and justifications (validation) of the relationships are given in the next section together with definitions of concepts.

### VALIDATION OF FRAMEWORK (CONTENT VALIDITY)

The paired quality and ergonomics concepts will be discussed based on the sequence in Table 36.1.



**FIGURE 36.1** Two areas of ergonomics and quality integration.



**TABLE 36.1**  
**Framework for Integrating Ergonomics with Quality**

| Quality vs. Ergonomics Pair <sup>a</sup>                                     | Complementary Area                                                                                                                                                                                                                                                                                                                                           | Contradictory Area                                                                                                                                                                                                                                                                               |
|------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                              | Easy Integration                                                                                                                                                                                                                                                                                                                                             | Difficult Integration                                                                                                                                                                                                                                                                            |
| 1. Total quality management (TQM) vs. socio-technical systems (STS) approach | <i>TQM Scandinavian/European &amp; Japanese traditions ✓ STS approach</i><br>Focus on people and processes: long-term goals with emphasis on teamwork, participation, working condition, culture, values and trust.                                                                                                                                          | <i>TQM American tradition ✗ STS approach</i><br>Focus on tools: cost-reduction tools, short-term results orientation and controlling (rather than empowering) employees.                                                                                                                         |
| 2. Customer focus (CF) vs. ergonomics stakeholder model                      | <i>CF ✓ Ergonomics Stakeholder Model</i><br>Customer satisfaction is achieved through employee motivation and involvement is consistent with giving consideration to all stakeholders.<br>Customers (in CF) include internal customers, i.e. employees.                                                                                                      | <i>CF ✗ Ergonomics Stakeholder Model</i><br>CF gives highest priority to customers in all situations, whereas Ergonomics Stakeholder Model considers all stakeholders in an equitable way.                                                                                                       |
| 3. Quality policy deployment (QPD) vs. participatory ergonomics (PE)         | <i>Participative QPD ✓ PE</i><br>Participative QPD includes employee participation in policy development, deployment and goal setting (including ergonomics goals).<br>Encourages motivation, learning, commitment and improvement.                                                                                                                          | <i>QPD ✗ PE</i><br>QPD presents no challenge to power structures, focuses on short-term profit and bureaucracy, whereas PE encourages employees' involvement in planning and controlling their work, with knowledge and power to influence processes and outcomes.                               |
| 4. ISO 9000 vs. psychosocial and organisation (PSO) factors                  | <i>ISO9000 developmental view ✓ PSO</i><br>Better working environment, participation, communication, learning and innovation; smoother working operations, clearer job description and responsibilities and improved work structure.<br>Job enrichment and enlargement.<br>Broader worker competence.                                                        | <i>ISO9000 regulatory and marketing views ✗ PSO</i><br>Less varied tasks, rule-based, monotonous, repetitive work, bureaucratic, meaningless tasks counteract learning, creativity and motivation; time pressure stress, increased responsibility without control and physically strenuous work. |
| 5. Process orientation (PO) vs. systems ergonomics (SE)                      | <i>PO ✓ SE</i><br>Closely related definition of PO and SE, i.e. humans interacting with technology, organisations and environment, with a goal in mind.<br>Both focus on design of tasks, workplace and environment.<br>SE takes the whole system (many processes) into consideration, together with the goals of multiple stakeholders and integrates E/HF. | <i>N.A.</i>                                                                                                                                                                                                                                                                                      |
| 6. Zero defects (ZD) vs. systems ergonomics (SE)                             | <i>ZD ✓ SE</i><br>Both concepts are pro-active, making things right from the beginning of the process or system design.<br>Both concepts focus on prevention of errors but from different approaches.                                                                                                                                                        | <i>N.A.</i>                                                                                                                                                                                                                                                                                      |

(Continued)

**TABLE 36.1 (Continued)**  
**Framework for Integrating Ergonomics with Quality**

| Quality vs. Ergonomics Pair <sup>a</sup>                         | Complementary Area                                                                                                                                                                                                                                                                                                                                                                                                                                  | Contradictory Area                                                                                                                                                                                                                                                                 |
|------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                  | Easy Integration                                                                                                                                                                                                                                                                                                                                                                                                                                    | Difficult Integration                                                                                                                                                                                                                                                              |
| 7. Quality value chain (QVC) vs. ergonomics value chain (EVC)    | <p><i>QVC ✓ EVC</i></p> <p>Both add value to a company's value chain.</p> <p>Ergonomics is part of product quality characteristics offered to customers.</p> <p>Ergonomics leads to an optimised system with high quality.</p> <p>Ergonomics adds ethical consumption value.</p>                                                                                                                                                                    | <i>N.A.</i>                                                                                                                                                                                                                                                                        |
| 8. Continuous improvement (CI) vs. participatory ergonomics (PE) | <p><i>CI ✓ PE</i></p> <p>Both have elements of participation, empowerment and motivation.</p> <p>Both encourage adaptive and developmental learning.</p> <p>Both evolved from expert to participatory approach.</p> <p>Both share several similar tools.</p> <p>Many studies have successfully integrated CI and PE.</p> <p>Both have similar problem-solving steps and were integrated in SIMPLE and WISE.</p>                                     | <i>N.A.</i>                                                                                                                                                                                                                                                                        |
| 9. Quality tools (QT) vs. ergonomics principles and tools (EPT)  | <p><i>QT ✓ EPT</i></p> <p>Both share several similar tools.</p> <p>Ergonomics' principles address the deficiency of quality tools which do not consider human limitations.</p> <p>Many studies show ergonomics principles applied to the work environment (e.g. noise, vibration, climate, lighting, chemicals and physical workload, and psychosocial and organisational factors) improve human performance and subsequently quality outcomes.</p> | <p><i>QT ✗ EPT</i></p> <p>Control chart, process capability study and just in time tools result in higher responsibility, time pressure and anxiety; tightly paced production, lack of control, loss of freedom and reduced motivation, creativity and learning opportunities.</p> |

*Notes:* ✓, complements; ✗, contradicts; vs., versus; N.A., not applicable; SIMPLE, select, interpret, measure, progress, learn and echo; WISE, work improvement in small enterprise.

<sup>a</sup> Paired quality and ergonomics concept which has both complementary and/or contradictory relationships.

## TQM VERSUS SOCIO-TECHNICAL SYSTEMS APPROACH

TQM is a holistic quality philosophy inspired by practitioners from Japan and the United States (Deming, 1986). It developed from the writings of its founders Edwards Deming, Joseph Juran and Kaoru Ishikawa. Its origin is statistical quality control in manufacturing, although this has developed into an integrated, systematic organisation-wide philosophy, emphasising organisational and management issues. Different authors have different views of the basic principles and foundations of TQM, but the following cornerstones are generally regarded as the most central: customer focus, process orientation, employee participation and CI. Different TQM traditions can be distinguished, including the Scandinavian/European (Bergman and Klefsjö, 2003), the Japanese (Imai, 1986; Monden, 1994) and the American (Shiba et al., 1993). In the Scandinavian/European and Japanese

traditions, much emphasis is given to teamwork and participation (Shiba et al., 1993; Golomski, 1994; Hackman and Wageman, 1995; Bergman and Klefsjö, 2003) and working conditions, which *complement* ergonomics' socio-technical systems (STS) approach. The STS approach is defined as a human-centred design approach in an organisation's work system, which differentiates the social (human) and technical subsystems and proposes a parallel design of the subsystems to achieve joint optimisation (Waterson, 2015). TQM believes in the importance of people for quality performance, and that companies that put emphasis on empowerment and personnel policies improve quality outcomes more than other companies (Agus, 2004; Tari and Sabater, 2006). Several aspects have been given interest, for example issues of culture, values and trust (Dahlgaard, 2002). Conversely, the American tradition is more result oriented, emphasises cost reduction with a short-term view and has been criticised for being used more as a management tool for rationalisation control (Redman and Mathews, 1998). It is more biased towards tools rather than human concerns, and thus *contradicts* the STS approach (Waterson, 2015). TQM is seen as controlling rather than empowering; in other words, intending to force employees to work harder without rewards (Fok et al., 2000).

### **CUSTOMER FOCUS VERSUS ERGONOMICS STAKEHOLDER MODEL**

In any strategic decisions, there are three main stakeholders, that is employers, employees and customers (Eklund, 1998a). Employers have a strong interest in profitability and productivity that contribute to the survival of an organisation. Employees have a strong interest in their working conditions including salary, rewards and job recognition, while customers have a strong interest in the quality of products and services, and the prices they pay. TQM's customer focus concept gives the highest priority to the customer satisfaction to increase profitability. Improvement of employees' working conditions (e.g. involvement, empowerment and job security) is seen as a side effect of having more money with increased profitability (Hendricks and Singhal, 1997; Ooi et al., 2006). However, too strong a focus on customers may lead to poor working conditions such as a high workload, stress and time pressure (Eklund, 1997). It has also been found that customer focus for employees in service jobs is negatively associated with employee job satisfaction (Kim et al., 2005). Further, Karia and Asaari (2006) found that customer focus did not contribute to job involvement, job or career satisfaction or organisational commitment. Customer focus sometimes contradicts the Ergonomics Stakeholder Model, which believes in optimising human–systems interactions (Eklund, 1998a; Hendrick, 2008; Dul and Neumann, 2009; Dul et al., 2012; Wilson, 2014a) and considers all stakeholders in an equitable way (instead of the highest priority given to customers in the customer focus concept). It also predicts that dissatisfaction from one or two stakeholders tends to create an unstable situation, which threatens the system in the long run, and may therefore risk the organisation's survival. Conversely, the two concepts sometimes complement each other as customer satisfaction in customer focus is achieved through employee motivation and involvement. This is consistent with giving reasonable consideration to the needs of all stakeholders in the Ergonomics Stakeholder Model. Customer focus believes in fulfilling the needs of internal customers, that is employees, which is again consistent with the Ergonomics Stakeholder Model.

### **QUALITY POLICY DEPLOYMENT VERSUS PARTICIPATORY ERGONOMICS**

TQM's quality policy deployment (QPD), which is also known as Hoshin Kanri (Akao, 2004), is a top-down approach where the top management's goals (in terms of quality, cost, delivery and employee-related targets) are systematically deployed through policies to lower-level management, and then to the shop floor. This is combined with a bottom-up approach involving worker participation through CIs ('daily management') and breakthrough strategies. QPD contradicts participation as it has been criticised for not challenging power structures, to be a way for management to manipulate employees and to focus on short-term profit and bureaucracy (Klein, 1984; Wilson et al., 1993; Hackman and Wageman, 1995; Drury, 1997). The way TQM approaches empowerment/

participation has been criticised as superficial and trivialising conflicts that exist within organisations (Wilkinson, 1998). Participation in the E/HF context (or PE) is defined as: 'The involvement of people in planning and controlling a significant amount of their own work activities with sufficient knowledge and power to influence both processes and outcomes in order to achieve desirable goals' (Wilson, 1991; see Chapter 34 of this book). Participation can take place at higher or lower levels, exemplified by 'user control' as the highest level, and 'consultation' or even 'user manipulation' as a negative form of participation (Arnstein, 1969). Among the advantages of participation are improved motivation and skills, personal development, better quality solutions, 'a value in itself' and legitimisation of worker experience. In general, participation is one of the most important key factors for sustainable change (O'Brien, 2002). Caccamise (1995) described how participative approaches lead to quality as well as ergonomics improvements. Many studies also point out that participation in ergonomics improvements results in better quality output (Moore and Garg, 1998; Lee et al., 2003). Axelsson proposed a model for participative QPD (which he termed as 'a generic model of quality and ergonomics management and improvement') in which employees participate in the policy development, deployment and goal setting (Axelsson, 2000, p. 357). This participative QPD approach also balances the ergonomics' goals (e.g. learning and growth) together with the company's other goals (e.g. financial, customer and internal business processes) with the use of a balance scorecard. In the deployment stage, the ergonomics' goals are used as input to the select, interpret, measure, progress, learn and echo (SIMPLE) process of CIs. This participative QPD approach encourages motivation, learning, commitment and improvement, and thus complements the PE (Axelsson, 2000).

### **ISO 9000 QUALITY MANAGEMENT SYSTEM VERSUS PSYCHOSOCIAL AND ORGANISATIONAL FACTORS**

ISO 9000 is a quality management system, which is considered a first step towards TQM (Bergman and Klefsjö, 2003) for an organisation. Its effects on operations have fallen short of expectations in many companies (Conti, 2004). One reason is that standardisation is often introduced with a marketing view, where companies only consider the market advantages when introducing ISO 9000, while the opportunity to obtain more effective processes is often overlooked (Karlton et al., 1998; Garvare, 2002; Poksinska et al., 2003). Another reason is that standardisation is introduced with a regulatory view instead of a developmental view (Eklund et al., 1998). With the developmental view, management introduces the standard with the major aim of improving internal processes. A participatory approach is applied, whereby the employees themselves document the processes for the standard. The standard is regarded as a temporary agreement until a better way to perform the work is found and standardised, and the standard is actively and continuously revised. The regulatory view is in many respects the opposite of the developmental view. An expert consultant is hired to write the documentation, which is handed out to the employees just before the consultant leaves the company. This documentation is used to discipline the employees. Changes of the standard are avoided (Eklund et al., 1998).

The term psychosocial was coined by Eric H. Erikson (Erikson, 1959). The psychosocial and organisational (PSO) factors are defined as psychological and social factors in the work environment in which an individual interacts with all components of the overall work situation (Bohgard et al., 2009, p. 23). Bongers et al. (1993) presented the factors in two categories, that is demand and control and social support. Demand and control include monotonous work, concentration, time pressure, responsibility, workload, work clarity, rest breaks and autonomy, while social support includes support from colleagues and superiors. If ISO 9000 is practised according to *regulatory and marketing views*, it contradicts the PSO factors of E/HF by creating less varied tasks, rule-based work, monotony, repetitive work injuries, and, in that way, counteracts learning, creativity and motivation. Monotonous and repetitive jobs give rise to symptoms of boredom or fatigue, followed by a decrease in performance in terms of longer reaction time and an increase in error rate progressively during the work period (Grandjean, 1988). Drury and Prabhu (1994) emphasise that humans have inherent difficulties in achieving satisfactory quality in repetitive

tasks at a high pace and for long periods. The other negative effects include increased bureaucracy with new, apparently meaningless tasks for individual workers (such as unnecessary paper work), as well as more stress in terms of time pressure, increased responsibilities without control and more physically strenuous work.

If ISO 9000 is practised according to the *developmental view*, it will complement E/HF as it enhances the PSO factors. In the developmental view, standardisation and self-discipline are necessary structures for worker participation, better communication, innovation, improvement activities and smooth working operations, in other words, to obtain quality, low costs, and safe and healthy working conditions (Imai, 1986; Kondo, 1995; Zuckerman, 1996). The standard should be seen as an important reference, and a basis from which learning and creative improvements can take place (Shiba et al., 1993; Kondo, 1995; Eklund et al., 1998). In addition, job descriptions and responsibilities become clearer, and job enrichment and enlargement result. Employees appreciate reduced role ambiguity and improved structure of their work (Karlton et al., 1998; Lundmark and Westelius, 2006). Drury and Prabhu (1994) point out that job enrichment improves inspection performance. Job enlargement is considered to contribute to better quality performance (Molleman et al., 2001). Operators with broader assembly competence have been shown to perform substantially better than those with narrower competence. Moreover, the latest ISO 9001 (2008) quality standard has a requirement on providing a suitable work environment, which includes E/HF aspects such as lightings, noise and temperature as well as psychosocial factors in the workplace (ISO 9000, 2008). This requirement complements E/HF if implemented according to the developmental view as a work environment is dynamic and ever-changing due to the short product and service life-cycles. E/HF aspects have to be constantly evaluated and improved, and there is no way they can be fixed or remain the same as with the regulatory view. For example, whenever there is any product/service changes, work processes have to be changed accordingly. If the regulatory view is followed, the important PSO factors may not be reviewed or addressed adequately according to the change, for example job enrichment and enlargement, challenges, learning, ownership and recognition, since the human resource policies and standards are 'cast in stone'.

## PROCESS ORIENTATION VERSUS SYSTEMS ERGONOMICS

The quality movement emphasises process orientation as one of its cornerstones, that is if the processes are running well, the result will subsequently be high-quality products and output. In process orientation, a process consists of human activities, where humans interact with technology and other humans within the organisation and environment, with the aim of transforming an input to a desired output (Melin, 1999). The quality of the human activities will decide the quality of the end product (Eklund and Bergman, 2003). Process orientation focuses on the design of tasks, workplaces and environment; however, it does not incorporate the E/HF knowledge in processes (Drury, 1997; Eklund, 1997). This limitation can be addressed by complementing it with the system ergonomics concept, which has a very similar but broader definition, that is a system with a set of processes consisting of human activities interacting with technology, organisations and environments, that transforms input to output for the benefit of multiple stakeholders in society with economic (e.g. quality, productivity and profit), social (e.g. well-being of workers) and environmental outcomes (Wilson, 2014a,b). The main difference is that systems ergonomics has a wider view, more encompassing goals and offers an opportunity for redesigning the overall work systems (rather than individual tasks, workplaces and environments in silos) to better fit humans, and therefore improve quality. It also integrates human factors and ergonomics (e.g. physical, cognitive and organisational ergonomics) into the complex systems (which is lacking in the process orientation approach). For example, Carayon et al. (2014) presented the systems engineering initiative for patient safety (SEIPS) model (an example of systems ergonomics) and claimed it has been successfully integrated with the process orientation concept (in TQM) as shown by its applications in many quality improvement studies in healthcare with human factors taken into consideration. It has

been applied in the follow-up process of outpatient treatment (Singh et al., 2009), evaluation of the safety of electronic health records (Sittig and Singh, 2009) and improvement of patient safety in radiotherapy (Rivera and Karsh, 2008).

### **ZERO DEFECTS VERSUS SYSTEMS ERGONOMICS**

One of the quality philosophies is zero defects, that is doing it right the first time through defect-prevention systems (Crosby, 1979). It is aimed at motivating employees to prevent errors and constantly desire to do the job right the first time. This concept was practised in Kaizen's CI where quality is built into the design of processes to prevent errors (Womack et al., 1990). Later, the concept was adopted to reduce work accidents in more than 280 Finnish industries, and named zero accident vision (Zwetsloot et al., 2013). The rationale was that zero defects could be achieved through prevention of accidents in heavy industries such as oil and gas as such accidents will lead to defects. The zero defects concept complements the new systems ergonomics paradigm (Wilson, 2014a,b) by optimising the interactions involved with the integration of systems elements (e.g. human, technical, information, social, political, economic and organisation). In this paradigm, working conditions should be seen as a proactive issue in a systems perspective, in integration with the demands on the organisation, such as high quality, flexibility, short lead times, delivery precision, undisturbed production, motivated employees and a good atmosphere of collaboration. Thus, systems ergonomics believes in setting things right from the beginning as it proactively integrates ergonomics into all decisions when planning technology (product and production design), working environments as well as work organisations. Both system ergonomics and zero defects focus on prevention of errors but use different approaches. System ergonomics takes a broader system design view than zero defects (Carayon et al., 2014). In addition, it explicitly incorporates human factors and ergonomics aspects while zero defects does not.

### **QUALITY VALUE CHAIN VERSUS ERGONOMICS VALUE CHAINS**

Quality is accepted as a company's function and part of its value chain as it creates value for the customer through quality products and services (Dul and Neumann, 2009; Dul et al., 2012). E/HF is not usually seen as part of the company's value chain. The old paradigm stated that ergonomics and working environment issues were often at the centre of conflict and bargaining between unions and management, and that ergonomics improvements were seen as a costly expenditure. Quality complements the new paradigm of 'ergonomics value chain', that is ergonomics adds value to a company's value chain as better working conditions/ergonomics lead to an optimised system with high quality, flexibility, short lead times, delivery precision, undisturbed production, motivated employees and a good atmosphere of collaboration (Dul and Neumann, 2009; Dul et al. 2012; Wilson, 2014a). In addition, some companies have a product strategy, which includes ergonomics properties (e.g. easy and safe to use) as part of product quality characteristics. Furthermore, selling products with ergonomic properties means that companies need to integrate good ergonomics in their production to maintain credibility, that is satisfying their internal customers (i.e. employees). This concept is called 'ethical consumption' where the value provided to the customers includes the assurance that an ergonomic/ethical working condition is practised in the production of the goods (Smith, 1990; Zander and Hamm, 2010).

### **CONTINUOUS IMPROVEMENT VERSUS PARTICIPATORY ERGONOMICS**

Continuous improvement (CI) is a bottom-up approach in TQM and can be used to improve product quality, production performance, safety or ergonomics (Eklund, 1998b; Fuller, 1999). Lillrank et al. (2001) proposed the following definition: 'Continuous Improvement is a purposeful and explicit set of principles, mechanisms and activities within an organisation adopted to generate continuous and systematic improvement in deliverables, operating procedures and systems by the people who actually



perform these procedures and operate the systems. CI contributes positively to the organisation's target achievement'. There are three key programmes of CI: employee suggestion schemes, quality circles (QCs) and the problem-solving cycle, all of which involve high worker participation. The employee suggestion scheme is a structured programme whereby the company rewards its employees for suggestions to improve their work with quality, productivity, cost-savings and occupational health and safety goals in mind. QC is the use of small group activities for work improvements. Representatives from management often choose themes for QC, the most common of which are cost, quality and efficiency. QC supports the PSO factors as it comprises of a set of methodologies that empowers employees to contribute solutions to work environment problems. It encourages participation that contributes to both individual and organisational learning, leading to good and rewarding work. The problem-solving cycle refers to the Deming PDCA cycle (plan, do, check, act) and the seven quality control steps: (1) select theme, (2) collect and analyse data, (3) analyse causes, (4) plan and implement solutions, (5) evaluate effects, (6) standardise solution and (7) reflect on the process. In many companies, the selection of themes is controlled and management makes decisions after proposals from a particular group. CI has progressed from being expert oriented to a participatory approach where in the initial version of the PDCA cycle (which was found inadequate), managers were planning, workers were doing, inspectors were checking and managers were taking action (Imai, 1986). When there is participation throughout the problem-solving cycle, a situation is created whereby learning is supported through reflection, analytical thinking and comparison with earlier experiences. In addition, when problem-solving has been carried out several times through the whole cycle, it provides opportunities for second-order learning, that is learning how to learn (e.g. developmental learning). Conversely, if the implementation of the solutions is left to specialists, learning will be impaired and take place at lower levels, for example adaptive learning. This learning process creates positive effects such as personal development, broader job content, better feedback, opportunities of influence, social contacts, variability and challenges. All these factors enhance motivation and ownership (Hackman and Oldham, 1980). The transition of an organisation to participative problem-solving means a move towards coherence with the characteristics of good work (Eklund, 1998b).

PE (defined in the earlier section, and discussed in detail in Chapter 34 of this volume) entails the involvement of workers in the improvement of their work and working conditions through the provision of the necessary knowledge, skills, training and empowerment to conduct such improvement. The rationale is that workers know best about their workplace and are in the best position to identify improvements. PE increases teamwork, motivation and learning in the process of these improvements. It is argued to be better than the traditional expert-oriented approach as it results in deeper learning among employees (developmental learning) compared with a low level of learning (adaptive learning) in the latter approach (Axelsson, 2000, p. 214; Wilson et al., 2005, p. 936). Chapter 34 presents a comprehensive set of methods, techniques and tools used in PE.

A strong complementary relationship exists between CI and PE. Participation and motivation are central concepts in both and, to some extent, other issues such as empowerment in the area of the psychosocial working environment. Both CI and PE encourage adaptive and developmental learning. They have also gone from an expert-oriented approach to a participatory approach. In addition, they have undergone some integration, focusing not only on organisational performance (e.g. quality) but also workers' well-being (i.e. working condition, motivation and second-order learning). CI has provided employees the opportunity to improve their working conditions; otherwise, there is a higher risk that it becomes a fad (Eklund, 1998b; Rapp and Eklund, 2007). Many industries have integrated both quality and ergonomics improvement in CI (Munck-Ulfsfält et al., 2003). For example, Axelsson (2000) found that 6 of 10 problem-solving projects within the CI programme in a large organisation were related to ergonomics. Also, Noro (1988) found a third of suggestions in employee suggestion schemes was related to ergonomics. Eklund (1999) also found from several Swedish cases that often 30%–50% of all suggestions in QC activities were related to ergonomics. Axelsson (2000) proposed the SIMPLE method that enhances the PDCA problem-solving cycle to include ergonomics and quality outcomes, and validated its effectiveness in a factory. He found the

method effective for quality and ergonomics improvements, participation and learning. Lee (2005) adapted the QC to address ergonomics goals, and thus developed a new PE method called ergonomics circles. He validated its effectiveness through a case study of a Korean electrical appliances factory. The PE approach is highly applicable to CI (Noro and Imada, 1991), and many studies have integrated PE with CI, for example Caroly et al. (2010), Kogi (2006, 2008), Axelsson (2000) and Haims and Carayon (1998). Noro and Imada (1991) presented 20 tools, while Kirwan and Ainsworth (2003) presented 25 tools that can serve both CI and PE. Some of the PE tools presented in Chapter 34 of this book such as Pareto analysis and cause and effect diagram are used in CI. The steps in PE of analysis, design, implementation and evaluation resemble the PDCA steps in CI. Due to their similarities, Axelsson (2000) and Kogi (2006, 2008) combined them to form the SIMPLE and work improvement in small enterprise (WISE) methods respectively.

### QUALITY TOOLS VERSUS ERGONOMICS PRINCIPLES AND TOOLS

TQM has a set of quality tools such as the seven quality control tools, which consist of the cause and effect diagram, check sheet, control chart, histogram, Pareto, scatter diagram and stratification (Tague, 2005). These quality tools evolved and extended to more than one hundred quality improvement tools including brainstorming, affinity diagram, arrow diagram, etc. most of which are for planning and implementation, idea creation, process analysis, data collection and analysis, cause analysis, and evaluation and decisions (Tague, 2005). Many of them complement E/HF as they are used to generate group participation, and are the same tools used in PE such as the checklist, brainstorming, sampling, interview, questionnaires, focus group, Pareto and cause and effect diagram and requirement analysis (Tague, 2005; Wilson et al., 2005). However, a few of these quality tools contradict ergonomics principles such as the control charts, process capability study and just in time (Drury, 1997). For example, Kelly and Drury (1999) identified social and technical reasons for the failure of control charts such as trust in the data, the reaction pain/stress of needing to take action in out-of-control situations, the competing objective of productivity, which supersedes the need to stop the process when such a situation arises, and the absence of technical help during night shifts, which limits operators' ability to get the process back under control. Klein (1989) found that the use of control charts results in the loss of individual and team autonomy as workers have to strictly adhere to the continuous monitoring of the process; stopping it to take action does not allow creativity to use other problem-solving methods. Process capability studies force operators to reduce the process variability to meet the process capability index. This reduces the decision-making autonomy of operators by forcing them to adhere to the 'one best way' of working (Drury, 1997). Again, this does not allow workers to be creative in attaining the process capability index through an alternative system redesign. In addition, the specification in the process capability index is set by the customer and has no relation to the process; thus, the process may not be able to meet the specification and this may cause workers' anxiety in trying to reach an impossible goal. Wilson et al. (1993) found that just in time creates a cellular manufacturing process, which involves tightly paced production that has adverse E/HF consequences to workers. The three tools mentioned add more responsibility, increase the pressure on work intensity and hamper motivation and learning with standardised and routine tasks (Klein, 1989; Wilson et al., 1993; Hackman and Wageman, 1995). Quality tools should be complemented with E/HF principles during quality improvement whenever the human element is involved. This will address the deficiency of quality tools, which do not take into account human limitations. Many studies have shown that environmental factors, such as noise, vibration, climate, lighting, chemicals and physical workload, and PSO factors set limitations on human performance, and subsequently affect quality outcomes (Sanders and McCormick, 1993; Helander, 2006). In such studies, it was found that when E/HF principles were used, quality improvements accelerated together with higher worker motivation and better working conditions (Eklund, 1999; Axelsson, 2000; Rapp and Eklund, 2002). Some useful E/HF principles for quality improvement are described in the following.

## Noise

Noise is defined as unwanted sound. It has a masking effect that impairs performance in situations where sound cues are used at work. For example, Axelsson (2000) showed that in the assembly of components with a snap-on function, the snap sound signalled 'passed' to the worker. When higher noise levels were introduced at the workplace (e.g. noisy work equipment), the snap sound was obscured. This led to a higher number of quality deficiencies. Masking threshold was developed to measure the hearing threshold of sound cues in a noisy environment, and it was suggested to be set at 75% probability of correct detection (Haslegrave, 2005). In addition, arousal is strongly related to performance. Arousal is increased by noise in frequencies that can be heard, but arousal is decreased by infrasound that has too low frequencies to be heard. Furthermore, noise has been shown to impose stress reactions, disturb concentration, cause distraction and a lapse of attention, all of which may influence the quality of work. For example, Banbury and Berry (2005) found 99% of office workers (respondents) lost concentration due to noise from telephones ringing and people talking in the background, and 59% indicated that sound-absorption walls and quieter places would improve performance. In several studies, noise has been shown to increase the error rate, particularly in tasks involving mental work (Smith-Jackson and Klein, 2009). For example, Taylor et al. (2004) found random and intermittent noise patterns can have a negative effect on the accuracy of human inspection. However, there are also studies that report improvements in performance due to sound in certain situations, where the character of the sound and type of work seem to be influential (Taylor et al., 2004). One example of this is music in monotonous assembly jobs.

## Vibrations

Hand-held vibrating equipment may impair finger sensitivity in the short term, and may also cause vibration white finger and other occupational disorders in the long term (Gold et al., 2005). Jobs in, for example, the wood industry often evaluate surface quality by touching and feeling the surfaces of manufactured products with hands and fingers (Karlton and Eklund, 1997). Difficulties in perceiving sensations with the fingers consequently cause poor performance and increase the risk of errors and accidents. Whole-body vibrations which lead to oscillations of the eye can be perceived as vibration of the viewed object and difficulty to see small objects. One such effect of vibration is that the error frequency in reading tasks may increase (Bhiwapurkar et al., 2011). In manual-control tasks, vibrations and jerks may cause errors due to unwanted movement of the controls. Threshold vibration frequency depends on contact conditions and vibration magnitudes; for example for a typical hand grip, frequencies above 125 Hz may interfere with the neuromuscular processes, including finger sensitivity, and thereby cause errors (Griffin, 2012).

## Climate

Body temperature influences many aspects of human performance, for example alertness, working memory and visual attention. Performance is improved at body temperatures near its circadian peak (Wright et al., 2002). Several studies have identified loss of grip strength, reduced manual performance and tactile sensitivity. For example, Chi et al. (2012) found that performance of fishermen and divers was affected by cold immersion, particularly in terms of their grip force. Goonetilleke and Hoffmann (2009) found that manual tracking performance (related to finger dexterity) reduced with low temperatures of about 10°C compared with 20°C and 30°C. Tactile sensitivity is important for manipulation of small objects, for example typing, assembly and small repair tasks (Goonetilleke and Hoffmann, 2009). Meese et al. (1980) reported a decrease in performance for manipulative tasks of approximately 10% when finger temperatures fall to 20°C due to reduction in tactile sensitivity. In a literature review, Seppanen et al. (2005) found a 2% decrease in work performance per degree centigrade when the temperature is above 25°C as high temperatures cause sweating, which not only distracts attention and affects concentration but might also result in a slippery hand grip.

## Lighting

Many studies have identified increased rates of misjudgement and inspection errors due to insufficient lighting. Many factors contribute to this, such as too low light levels, insufficient colour rendering, non-optimal luminance and disturbing reflections (Grandjean, 1988; Zegers and van den Berg, 1988). The same adverse effects have been observed in tasks such as proof-reading of texts with poor visibility (Dillon, 1992). Visual demands that put high stress on any of the functions of the eye can lead to poor quality output as in many documented cases from production industries where increased illumination levels resulted in substantial reductions in rejection rates (Grandjean, 1988). For example Chi et al. (2012) found that performance Helander and Burri (1995) showed that improvements in illumination resulted in a higher detection rate of faulty items. In an overview of 15 published studies, Juslén and Tenner (2005) reported a reduction of rejects, often varying between 10% and 60% after an increase in illumination levels. Howarth (2005) posited that specular reflection and glare have a negative effect on visual performance as they cause annoyance to the viewer. Yeow and Sen (2004) confirmed this through a field experiment in an electronic motherboard factory via a visual ergonomics intervention that changed the angle of the inspection template to avoid specular reflection in the inspection operator's line of sight and replaced the existing template with less reflection. This intervention resulted in lower rejection rates, better quality for customers and better visual comfort for operators. In a corresponding way, substantial increases in the number of defects found in inspection tasks have been seen after changes in the light source and angle. For example, surface irregularities, such as dents, wave shapes or defects in plane surfaces, can be easily detected with a special light source giving light at a low angle to the surface. Inspection of the smoothness of painted surfaces improves if light is directed so that the inspector can view the reflection. Surface flaws can be easier judged in diffused light from sources with large light-emitting areas or from indirect light (Schneider et al., 2013). Inspection of colour deviations is aided by light sources with a high colour-rendering index in combination with daylight properties (Boyce, 2003).

## Chemical Health Hazards

Solvents and lead are two examples of chemicals with a potential of causing a wide range of symptoms through effects on the nervous system, including nausea, dizziness, headache, concentration difficulties, memory loss, balance disturbances, longer reaction time, impaired dexterity and even personality changes such as sudden aggressiveness (Stollery, 1992; Akila et al., 2006). Other studies have identified lower performance in visual perception, colour vision, vigilance and psychomotor functions due to solvent exposure (Viaene, 2002). The use of protective equipment is a secondary effect due to deficiencies in working conditions. In general, such equipment may have many adverse effects such as being heavy or uncomfortable to use, making natural movements or activities more difficult, and impairing perception, communication and information retrieval. All these aspects influence human performance and accuracy. In many cases, gloves are worn for protection against health hazards such as chemicals, low temperatures or cuts. Consequently, precision decreases (Cushman and Rosenberg, 1991) and tactile feedback deteriorates (Sanders and McCormick, 1993; Shih et al., 2001).

## Physical Workload

The precision of body movements varies with the direction of movement, and also between different muscles in the body; hence, ergonomics design that takes this into consideration will enable greater accuracy of performance (Grandjean, 1988). Heavy and strenuous work tasks and awkward postures often result in increasing discomfort during the working day, in addition to increasing the risks of musculoskeletal disorders (Corlett and Bishop, 1976). In such situations, performing the work itself, particularly when making additional corrections and improvements of the work result, becomes even more strenuous and possibly painful. Thus, the willingness to make corrections and improvements to work of borderline quality diminishes, and quality deficiencies increase with the increase in discomfort level (Eklund, 1995). Several studies have identified a relationship between quality deficiencies and ergonomically problematic work tasks to the extent that about 30%–50% of

all quality deficiencies are related to or directly due to ergonomics problems (Eklund, 1995, 1999). For example, Falck et al.'s (2014) empirical study found that a high ergonomics (physical) load level in areas of workload, work posture, material handling and tooling has an effect on car assembly failure rate. In addition, Yeow and Sen (2003) refer to E/HF improvements in a printed circuit assembly plant, including workstation layout and posture improvements that resulted in decreased rejection rates, and improved quality and customer satisfaction. In another study, Lin et al. (2001) measured the number of defects at each workstation on two assembly lines. Two factors, that is the time required for the task and the postural deficiencies, together accounted for over 50% of the quality variance on each assembly line. Axelsson (2000) used the rapid upper limb assessment (RULA) tool to classify the appropriateness of upper-limb work postures (adapted from McAtamney and Corlett, 1993), and compared this measure with the quality outcome. The quality deficiency rate was almost 10 times higher for the worst posture compared with the best posture. His study shows the effectiveness of combined tools from different disciplines as the RULA ergonomics tool was used together with quality tools such as the Pareto analysis, scatter diagram and cause and effect diagram. Gonzales et al. (2003) also found a relationship between improved workplace design and better work posture and better quality in terms of less rework. Furthermore, Yeow and Sen's (2006) study on manual assembly concludes that unobstructed accessibility, which allows better work postures, is associated with better quality in terms of correct attachment/insertion of parts.

### **Psychosocial and Organisational Factors**

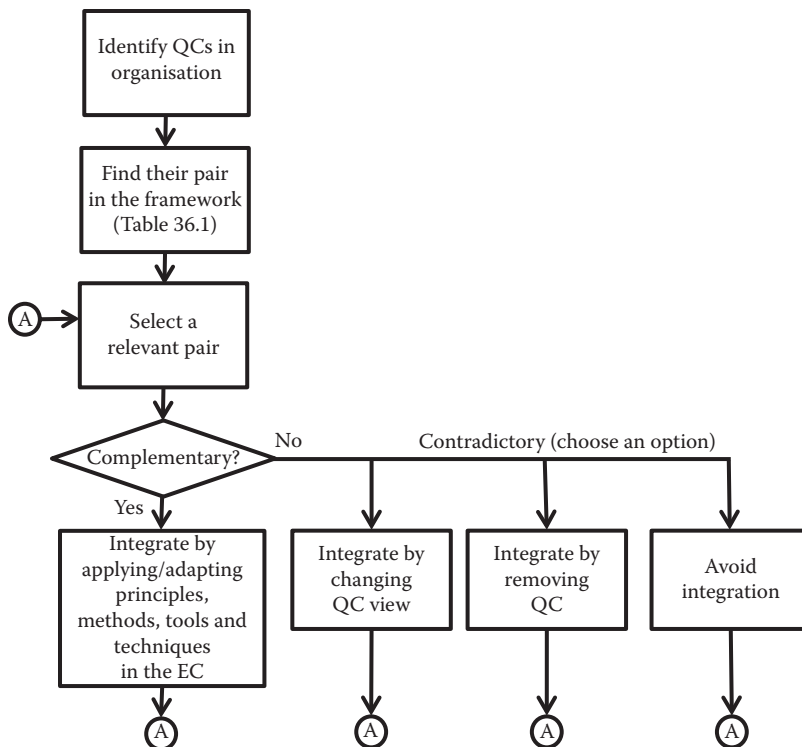
Self-paced work in relation to machine-paced work has been shown to improve quality (Gallwey, 1998). However, there are reports that job rotation may lead to decreased product quality (Jorgensen et al., 2005). Status differences and tensions between categories of workers in a hierarchical work organisation have also been related to quality deficiencies (Eklund, 1995). Several studies have documented that production philosophy, the work organisation and personnel policy as well as wage form may influence quality (Deming, 1986; Womack et al., 1990). In one study, it was found that U-shaped assembly lines resulted in better quality performance than straight lines (Cheng et al., 2000). It has been reported many times that motivation and job satisfaction lead to improved product quality (Lammermeyr, 1990). Among the important factors for this are a continuous interest in the work, challenges, recognition for achievement, ownership and communication. However, the generality of the assumption that job satisfaction automatically leads to improved quality has been challenged (Engström et al., 1996).

Another important influence on quality is the work organisation. Drury and Prabhu (1994) found that job enrichment improved quality performance, and data from Eklund (1995) showed improved assembly quality when self-inspection was included in the assembly tasks compared with using separate quality inspectors. In addition, Faye and Falzon's (2009) study found that operators' self-monitoring (a type of job enrichment) supports quality improvement in automotive production. Operators with broader assembly competence were found to perform better compared with those with less competence (Lovén and Helander, 1997). Furthermore, performance feedback to workers provides useful information, motivating effects and reinforcement of correct behaviour, thus improving sensitivity of the operators to produce higher quality performance (Annett, 1966; Seong and Bisantz, 2008; Goomas and Yeow, 2010).

### **GUIDELINES TO USE THE FRAMEWORK**

The key objective of this chapter is to provide E/HF practitioners with ideas to integrate ergonomics with quality and to use quality as argument to promote E/HF. The framework presented in Table 36.1 can be used to support integration of a range of E/HF concepts. It presents nine common quality concepts widely practised worldwide over the past few decades, that is TQM, customer focus, QPD, ISO9000, process orientation, zero defects, quality value chain, CI and quality tools. Each of these concepts is matched with the closest complementary/contradictory ergonomics concept.





**FIGURE 36.2** Integration steps. (Note: QC, quality concept; EC, ergonomics concept.)

A step-by-step approach for using the framework to integrate ergonomics and quality concepts in an organisation is presented in a flowchart (Figure 36.2). Firstly, the ergonomist/ergonomics practitioner identifies the quality concepts practised in his/her organisation. Secondly, for each of the quality concept, he/she refers to the framework (Table 36.1) to find its pair (of complementary/contradictory quality and ergonomics concepts). Next, he/she selects a relevant complementary/contradictory pair depending on his/her involvement in the company's quality initiative and ergonomics role. For example an ergonomist may be an adviser to a QC group and select the CI and PE pair (Table 36.1, No. 8), since he/she is involved with the CI initiative and has a role in improving the ergonomics conditions of the organisation. Next, if the pair is complementary, he/she applies/adapts the ergonomics concept, including the principles, methods, tools and techniques to the quality concept. For example the ergonomist (in the prior example) may attempt to apply the principles, methods, techniques and tools in PE (Wilson et al., 2005) to the QCs. He/she could put emphasis on applying those methods with prior success, for example WISE (Kogi, 2006, 2008) and SIMPLE (Axelsson, 2000). Finally, if the pair is contradictory, he/she could (1) attempt to change the view of the quality concept so as to complement ergonomics concepts, for example changing from an ISO 9000 regulatory or marketing view to a developmental view to complement the ergonomics concept on PSO factors. Although this may seem like just a change in the quality concept view, it should be regarded as ergonomics integration with quality, because the change to a developmental view will address/improve the many ergonomics concerns in the regulatory and marketing views. Alternatively, he/she could (2) choose to remove the contradictory concepts, for example removing the use of control charts, process capability study and just in time to prevent ergonomics problems relating to time pressure, anxiety, tightly-paced production, etc. This example should also be regarded as ergonomics integrated with quality as the decision to remove quality tools that cause bad human working conditions is based on an analysis of the quality tools using ergonomics



principles. He/she could also (3) choose to avoid integration, since the quality concept is contradictory. This option should be chosen if he/she is powerless to influence the change of view or removal of the contradictory concept.

## CASE STUDIES

The purpose of the case studies is to illustrate integration/failure of integration based on the complementary/contradictory relationship in the framework. This will provide practitioners real-life and detailed examples on how to integrate ergonomics or how failure of integration occurs. Due to space limitations, only two pairs from the framework were chosen. The first case study also provides an example on the step-by-step approach to use the framework.

### CASE STUDY 1: CONTINUOUS IMPROVEMENTS VERSUS PARTICIPATORY ERGONOMICS

A medium-size manufacturing company introduced a system for CI, which was followed by a group of researchers over a 10-year period, described in the following text and based on the following publications: Eklund (1998b), Axelsson (2000), Rapp and Eklund (2002, 2007). The aim of the system was to improve operations through incremental changes and support employee participation in the improvement activities. Suggestions were encouraged in many different ways. For example the quality manager responsible for the system in the beginning walked around the shop floor and asked employees for any problems they had seen. With the problem as a basis, he encouraged them to find a solution to the problem. He helped those who had problems and solutions to fill in a form in order to document the problem and proposed solution, and this form was submitted to the suggestion system. Others filled in suggestion forms spontaneously, individually or in groups.

The management wanted the system designed according to the following factors:

- Stimulate high participation rates among employees.
- Aim for a disturbance-free production through all types of suggestions.
- Encourage small-step improvements.
- Systematic documentation of the suggestions.
- See the system as a joyful thing.
- Feedback on suggestions within 48 h.
- Every suggestion received a small incentive (a ticket to the cinema).
- An incentive for each group member for suggestions submitted by a group.
- Quick implementation.

The reason for encouraging small steps was so that it would be easy to submit suggestions, especially when the system was new. The experience showed that the complexity of the suggestions increased over the years. The employees learnt to fill in the form better. The suggestions became increasingly more advanced, more innovative and from more of a systems approach. In other words, the employees learnt to give better and better suggestions.

Rewards were given in order to make everyone aware of the suggestion system, to emphasise its element of fun and to increase participation. Through the system, participation rates could be monitored continuously, and reached around 80% for certain years.

Quick feedback on whether the suggestion was approved or not was also considered crucial for high participation and more suggestions. Also, the system of allowing any type of suggestions regarding quality, productivity, work environment, hygiene or environmental improvements contributed to a high interest in the system. There was also an effort towards quick implementation, which is another important form of quick feedback.

Participant incentives were used for all submitted suggestions. One reason for this was to give recognition to the employees. However, the incentives were small, according to the philosophy of

the system. In addition to the cinema ticket that was given to everyone who was part of submitting a proposal, small monetary rewards were given for accepted and implemented suggestions. Not all suggestions were approved, but over the years, between 50% and 70% of them were accepted. In fact, an analysis of the submissions showed that intermittent rejections increased the willingness to submit suggestions compared with if almost all submissions were accepted or rejected.

Group proposals were encouraged as the management felt groups would discuss the proposals internally and that the quality of suggestions would consequently improve. In this way, employees would collaborate in designing solutions and the sense of ownership would improve. Also, solutions would be accepted by all members of the group, which would create a broader sense of ownership, making implementation easier and solutions more sustainable. Individual rewards would have hampered group collaboration. The system was not misused by a large number of irrelevant names being signed to a suggestion just to obtain the reward.

The system was sustained for about 10 years, until the company was purchased by a competitor and merged, resulting in the closure of the site. The company succeeded in obtaining a very high participation rate (around 80%), with three to four suggestions per employee per year, increasing the number of group suggestions, and with suggestions becoming more advanced and innovative over time. These showed the success of the integration between ergonomics and quality.

This case may exemplify the three integration steps (Figure 36.2). First, the quality concepts practised by the company were listed in which suggestion system was one of them. Second, the pairs for the quality concepts were identified and finally the most relevant pair, that is CI and PE, was selected (Table 36.1, No. 8). The pair relationship was complementary and therefore integrated by applying the principles in the framework such as participation, empowerment, motivation, developmental learning and participatory approach. The case study showed how the quality manager encouraged participation from the employees, helped and guided them to find solutions to their work problems and to fill in the suggestion form and motivated them through small incentives. The suggestion system encouraged developmental learning as the employees learned how to solve problems by themselves and eventually increased the number of suggestions. The strong site manager support of the improvements empowered the employees to make changes to their work through the suggestion system.

In short, the system was basically a quality system for CI, but it made use of many concepts from the ergonomics' discipline. It aimed for high participation rates, using rewards and recognition, group work and management support. A key factor is that it made it possible for employees to improve their own working conditions. All these E/HF aspects supported participation as well as the function of the CI system, which was thus well integrated with the PE concept, making the systems mutually supportive.

## **CASE STUDY 2: ISO 9000 VERSUS PSYCHOSOCIAL AND ORGANISATIONAL FACTORS**

In a survey of manufacturing companies, different ways of applying ISO 9000 were seen. The analysis of these showed two principally different ways of applying and implementing ISO 9000. Most companies used a combination of the two. In this case, Company A represents a developmental view of standardisation, while Company B represents a regulatory and marketing view (Eklund et al., 1998; Karlton et al., 1998).

Company A: When the plant manager initiated the work with ISO 9000, he had a discussion with union representatives on the formation of a partnership around the change. The basic idea behind the standardisation was to improve processes in production. In the first phase, small work groups were formed on the shop floor, based on the existing production groups, where everyone was invited to participate. These groups were tasked with going through their processes and identifying which worked and which did not. They received support in their work. The processes that had problems with quality, productivity, disturbances or working environment were focused on, and a change process was initiated so that they could function better. Since this included technology, education, training and organisational changes, the first phase took over 1 year to implement. When the work groups were satisfied with the new processes, they moved to the second phase.

In the second phase, they were tasked with documenting their own processes and activities. Here too they received support on how to structure and write the documentation, but the responsibility was still within the work groups. In this way, the groups already formed a consensus among the members about the processes, which they considered best for their production. The work group displayed the process documents at the workplaces for everyone to see. The quality manager was given an important role in coordinating the groups, to support them and develop the standardised quality system.

The plant manager regarded the work of improving the processes and developing standardised processes (to get a structure and good quality processes in production) as the important aspect. The official certificate itself was considered of little importance.

In order to maintain interest in the standard, the plant manager had a conscious strategy. When he got a question from employees and managers related to production, he responded with the question: 'What does our quality system say about this?' Another cornerstone of the system was that the standard was regarded as a joint and temporary agreement, 'the best way that we presently consider we can perform our work'. At the same time, it was recognised that a better way will be found sooner or later when the standard will be revised; a joint and temporary agreement will contribute to a standard that is alive and improved continuously. In summary, ISO 9000 in this case included CI, participation, group work, motivation, responsibility, authority and a long-term developmental view. The standard became a sustainable tool in the company. Employees expressed high satisfaction with the ISO 9000 standard and conformed to the joint temporary agreement.

Company B: For market reasons, the plant manager wanted the official certificate as quickly as possible. There were also disciplinary problems he wished to solve through the standard. The first step was to employ an external consultant to write the standard. After some months of interviews and observation of the workers in order not to disturb them unnecessarily, the consultant wrote an extensive collection of documents on the standard. This was presented at a meeting with the personnel, and all workers were given a copy of the standard together with instructions to follow the prescribed processes. The company obtained its goal, that is the certificate, which was placed in the company reception, visible for all visitors. The plant manager had little use of the standard in the production processes, so it was mainly used as a rule for settling disagreements and maintaining discipline. In summary, the standard was in this case used as a regulation and not revised. It was imposed on the workers, making their work more difficult. Ergonomics concepts were not employed. The employees expressed strong dissatisfaction with the standard and resisted it. The standard created conflicts and did not become sustainable.

## COMPARISON WITH OTHER ERGONOMICS AND QUALITY INTEGRATION METHODS, FRAMEWORKS AND MODELS

Many studies have proposed methods, frameworks and models to integrate ergonomics with quality. The main sources of these studies are three books, two of which are conference proceedings of the *International Conference on TQM and Human Factors – Towards Successful Integration*, Volumes 1 and 2 Axelsson et al. (1999) and Axelsson's (2000) PhD thesis entitled *Quality and Ergonomics – Towards Successful Integration*. The conference proceedings consist of about 150 studies, while the thesis consists of 7 studies. The authors of this chapter also contributed to several other studies, that is Eklund (1995, 1997, 1999), Yeow and Sen (2003, 2006) and Erdinc and Yeow (2011). The present framework addresses the following limitations of the existing methods, frameworks and models:

1. None of the methods, frameworks and models provide a step-by-step integration of ergonomics concepts into a comprehensive set of the organisation's quality concepts. For example, Axelsson's (2000) thesis presents a generic model of quality and ergonomics integration, the SIMPLE method, and the integration of RULA and scatter diagram (one of the quality tools). However, his thesis has limited coverage of quality concepts. Erdinc and Yeow (2011) present five case studies of ergonomics and quality integration. However, they do not provide a

step-by-step approach to integrate the ergonomics concept with the quality concept. The 150 studies in the conference proceedings do not provide step-by-step integration guidelines. In addition, each of these studies involved the integration of a limited number of ergonomics and quality concepts. The present framework provides a wide coverage as it is based on 40 years of literature including step-by-step integrations of the major quality and ergonomics concepts.

2. None of them were developed based on complementary and contradictory relationships between ergonomics and quality concepts. Although some prior studies mentioned the complementary and contradictory nature of ergonomics and quality, their integration framework/model/method was not derived based on this (Wilson et al., 1993; Drury, 1997; Eklund, 1997; Axelsson et al., 1999; Axelsson, 2000). The complementary relationships in the present framework are useful as they identify compatibilities and similarities in the two concepts, which are common grounds for ergonomics and quality practitioners to initiate integration. Besides, the benefits of integration in terms of better working conditions and quality outcomes are attractive as proven (validated) in prior literature. The contradictory relationships in the framework are also useful for better integration decision-making as three options are provided, that is to change the view of the quality concept, remove the contradictory concept or avoid integration.
3. None of them proposed the change of the quality concept view as an integration option if there is a contradictory relationship. The present framework provides an alternative quality concept view with a complementary relationship for every quality concept with a contradictory relationship. This is a unique and useful characteristic, which is not available in other frameworks. For example, if the ergonomist discovers that his/her company is practising the customer focus concept and emphasising too much on external customers to the extent that working conditions are jeopardised, he/she may opt to change the view of the customer focus concept to emphasising internal customers instead and meeting the needs of employees to achieve quality outcomes.
4. Most of them only mentioned the effects of ergonomics on quality without explicitly providing an approach for matching and integrating the ergonomics concept with the quality concepts. For example, the SEIPS approach, classified under systems ergonomics, did not explicitly propose how to integrate it with the process orientation concept (Carayon et al., 2014). The present framework proposes integration through compatibilities and similarities between the process orientation and systems ergonomics concepts; it recommends the use of the systems ergonomics approach (e.g. SEIPS) to extend the process orientation concept to cover the human factors and ergonomics aspects, overall system (instead of limiting to a process) and goals of multiple stakeholders. In other examples, Eklund (1995, 1999), Helander and Burri (1995), Yeow and Sen (2003, 2006) and Gonzales et al. (2003) found quality improvement through the ergonomics approach; however, their studies did not explicitly provide an integration method. The present framework provides an approach for matching and integrating ergonomics with quality.

## LIMITATIONS OF THE FRAMEWORK

This section discusses the limitations of the proposed framework:

1. Exclusion of product-design concepts  
Due to space limitations, the framework did not include product design concepts such as affective design, quality function deployment, design for manufacturability, design for assembly and design for quality, design for ergonomics and Poka-yoke method, hedonic ergonomics, system analysis and design, and evaluation in human-computer interaction.
2. Overlap of concepts  
Some quality concepts in the framework are overlapping. For example the TQM views overlap, since the QPD, CF, ISO9000, PO, QVC, CI and QT concepts are part of TQM. However, these concepts can be differentiated through different management levels, that

is TQM views and QPD at the strategic level; CF, ISO9000, PO and QVC at the tactical level and CI and QT at the operational level. Similarly, there are overlaps in ergonomics concepts, which can be similarly differentiated, that is STS and participatory QPD at the strategic level; Ergonomics Stakeholder Model, PSO factors, SE and ergonomics value chain (EVC) at the tactical and PE tools and ergonomics principles and tools (EPT) (with the exception of PSO factors) at the operational.

3. May not be one-to-one relationships

The framework presents one quality concept linked with one ergonomics concept. However, since there are overlaps of concepts (as mentioned previously), some quality concepts may link to more than one ergonomics concept. For example, other than PE, CI may be linked to PSO factors as CI promotes continuous interest in work, provides challenges, recognises achievement and encourages ownership and communications.

Despite the limitations, the framework is useful as a starting point for integration. The main intention of the chapter is not to provide an exhaustive list of quality and ergonomics concepts and their match/matches, but to introduce the idea of using complementary and contradictory relationships as a basis for integration of the two concepts, and to provide the major relationships between the two concepts (i.e. the closest match of concepts). We recommend that practitioners/researchers expand the investigation to find more matches of concepts so as to customise and extend the framework specifically for their organisation's quality and ergonomics functions.

## CONCLUSION

The chapter presents an overview of the relationship between ergonomics and quality, and a framework to integrate ergonomics concepts with quality concepts. The overview shows strong mutual relationships between ergonomics and quality, and that ergonomics deficiencies lead to quality deficiencies. Further, quality improvements often lead to better work conditions and better ergonomics. The rationale for integration of ergonomics and quality in applied work is that there are many compatibilities and similarities between the two knowledge fields as clearly shown in the framework, where there is a complementary ergonomics concept for every quality concept. The framework consists of nine pairs of quality and ergonomics concepts, either complementary, contradictory or both (depending on perspective), and it is validated through prior literature. Guidelines to use the framework are given in a flow chart together with descriptions and examples. Two case studies of integration are also presented. The framework was compared with other existing integration frameworks, models and methods, and found to have the advantage of comprehensive coverage of concepts, a step-by-step approach, development based on complementary and contradictory relationships, an integration option to change the quality concept view in the case of a contradictory relationship, and matching and integration of quality and ergonomics concepts. The framework offers a new perspective for quality and ergonomics practitioners as a practical guideline to integrate ergonomics with an organisation's quality functions. It is also useful for researchers to identify the ergonomics and quality relationships that have been studied, and to research into new relationships so as to update the framework in the future. This is much needed as ergonomics and quality concepts are continually evolving according to the needs of different organisations and their stakeholders.

## REFERENCES

- Agus, A. (2004). TQM as a focus for improving overall service performance and customer satisfaction: An empirical study on a public service sector in Malaysia. *Total Quality Management and Business Excellence*, **15**, 615–628.
- Akao, Y. (2004). *Hoshin Kanri* (Portland, OR: Productivity Press).
- Akila, R., Müller, K., Kaukiainen, A. and Sainio, M. (2006). Memory performance profile in occupational chronic solvent encephalopathy suggests working memory dysfunction. *Journal of Clinical and Experimental Neuropsychology*, **28**, 1307–1326.



- Annett, J. (1966). Training for perceptual skills. *Ergonomics*, **9**, 459–468.
- Arnstein, S. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners*, **XXXV**, 216–224.
- Axelsson, J.R.C. (2000). Quality and ergonomics – towards successful integration [doctoral thesis] (Linköping, Sweden: Linköping University).
- Axelsson, J.R.C., Bergman, B. and Eklund, J. (eds.) (1999). The main sources of these studies are three books, two of which are conference. *Proceedings of the International Conference on TQM and Human Factors – Towards Successful Integration*, Volumes 1 and 2, 15–17 June 1999, Linköping, Sweden, vols. 1–2 (Linköping, Sweden: CMTO).
- Banbury, S.P. and Berry, D.C. (2005). Office noise and employee concentration: Identifying causes of disruption and potential improvements. *Ergonomics*, **48**, 25–37.
- Bergman, B. and Klefsjö, B. (2003). *Quality from Customer Needs to Customer Satisfaction* (Lund, Sweden: Studentlitteratur).
- Bhiwapurkar, M.K., Saran, V.H. and Harsha, S.P. (2011). Objective and subjective responses of seated subjects while reading Hindi newspaper under multi axis whole-body vibration. *International Journal of Industrial Ergonomics*, **41**, 625–633.
- Bohgard, M., Karlsson, S., Lovén, E., Mårtensson, L., Osvalder, A.L., Rose, L. and Ulfvengren, P. (2009). *Work and Technology on Human Terms* (Stockholm, Sweden: Prevent).
- Bongers, P.M., de Winter, C.R., Kompier, M.A. and Hildebrandt, V.H. (1993). Psychosocial factors at work and musculoskeletal disease. *Scandinavian Journal of Work, Environment and Health*, **19**(5), 297–312.
- Boyce, B.R. (2003). *Human Factors in Lighting* (London, U.K.: Taylor & Francis Group).
- Caccamise, D.J. (1995). Implementation of a team approach to nuclear criticality safety: The use of participatory methods in macroergonomics. *International Journal of Industrial Ergonomics*, **15**(5), 397–409.
- Carayon, P., Sainfort, F. and Smith, M.J. (1999). Macroergonomics and total quality management: How to improve quality of working life? *International Journal of Occupational Safety and Ergonomics*, **5**(2), 303–334.
- Carayon, P., Wetterneck, T.B., Rivera-Rodriguez, A.J., Hundt, A.S., Hoonakker, P., Holden, R. and Gurses, A.P. (2014). Human factors systems approach to healthcare quality and patient safety. *Applied Ergonomics*, **45**(1), 14–25.
- Caroly, S., Coutarel, F., Landry, A. and Mary-Cheray, I. (2010). Sustainable MSD prevention: Management for continuous improvement between prevention and production. Ergonomic intervention in two assembly line companies. *Applied Ergonomics*, **41**, 591–599.
- Cheng, C.H., Miltenburg, J. and Motwani, J. (2000). The effect of straight and U-shaped lines on quality. *IEEE Transactions on Engineering Management*, **47**(3), 321–334.
- Chi, C.F., Shih, Y.C. and Chen, W.L. (2012). Effect of cold immersion on grip force, EMG, and thermal discomfort. *International Journal of Industrial Ergonomics*, **42**, 113–121.
- Conti, T. (2004). How to conceptually harmonize ISO 9000 certification, levels of excellence recognition and real improvement. *Total Quality Management and Business Excellence*, **15**(5–6), 665–677.
- Corlett, E.N. and Bishop, R.P. (1976). A technique for assessing postural discomfort. *Ergonomics*, **19**, 175–182.
- Crosby, P.B. (1979). *Quality Is Free* (New York: McGraw-Hill).
- Cushman, W.H. and Rosenberg, D.J. (1991). *Human Factors in Product Design* (Amsterdam, the Netherlands: Elsevier Science Publishers BV).
- Dahlgaard, S.M.P. (2002). The human dimension in TQM: Learning, training and motivation [doctoral thesis] (Linköping, Sweden: Linköping University).
- Deming, W.E. (1986). *Out of the Crisis* (Cambridge, U.K.: Cambridge University Press).
- Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, **35**(10), 1297–1326.
- Drury, C.G. (1997). The Ergonomics Society Lecture 1996: Ergonomics and the quality movement. *Ergonomics*, **40**(3), 249–264.
- Drury, C.G. and Prabhu, P.V. (1994). Human factors in test and inspection. In W. Karwowski (ed.), *Design of Work and Development of Personnel in Advanced Manufacturing* (New York: John Wiley & Sons, Inc.), p. 355.
- Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W.S. and van der Doelen, B. (2012). A strategy for human factors/ergonomics: Developing the discipline and profession. *Ergonomics*, **55**(4), 377–395.
- Dul, J. and Neumann, W.P. (2009). Ergonomics contributions to company strategies. *Applied Ergonomics*, **40**, 745–752.



- Eklund, J. (1995). Relationships between ergonomics and quality in assembly work. *Applied Ergonomics*, **26**(1), 15–20.
- Eklund, J. (1997). Ergonomics, quality and continuous improvement – conceptual and empirical relationships in an industrial context. *Ergonomics*, **40**, 982–1001.
- Eklund, J. (1998a). Work conditions and company strategies. In P. Vink, E. Koningsveld and S. Dhondt (eds.), *Human Factors in Organizational Design and Management* (Amsterdam, the Netherlands: North-Holland), pp. 263–268.
- Eklund, J. (1998b). Participative problem solving and organizational congruence. In H. Boer and J. Gieskes (eds.), *The Second International EuroCInet Conference on Continuous Improvement: From Idea to Reality*, 14–15 September 1998, Enschede, the Netherlands, pp. 112–119.
- Eklund, J.A.E. (1999). Ergonomics and quality management – humans in interaction with technology, work environment, and organization. *International Journal of Occupational Safety and Ergonomics*, **5**(3), 143–160.
- Eklund, J. and Bergman, B. (2003). Developing work and quality improvement strategies – an introductory. *AI and Society*, **17**(2), 65–70.
- Eklund, J., Ellström, P.-E. and Karlton, J. (1998). Standardisation – a means for creating developing work? In *The First World Congress on Ergonomics for Global Quality and Productivity*, 8–11 July 1998, Hong Kong, China, pp. 165–168.
- Elg, M., Gremyr, I., Hellström, A. and Witell, L. (2011). The role of quality managers in contemporary organisations. *Total Quality Management & Business Excellence*, **22**(8), 795–806.
- Engström, T., Jonsson, D. and Medbo, L. (1996). The Volvo Uddevalla plant: Production principals, work organization, human resources and performance aspect (Gothenburg, Sweden: Department of Transportation and Logistics, Chalmers University of Technology). Report on work environment fund projects 93-0217 and 94-0516.
- Erdinc, O. and Yeow, P.H.P. (2011). Proving external validity of ergonomics and quality relationship through review of real-world case studies. *International Journal of Production Research*, **49**(4), 949–962.
- Erikson, E.H. (1959). *Identity and the Life Cycle: Psychological Issues Monograph*, vol. 1 (New York: International Universities Press).
- Falck, A.C., Örtengren, R. and Rosenqvist, M. (2014). Assembly failures and action cost in relation to complexity level and assembly ergonomics in manual assembly (part 2). *International Journal of Industrial Ergonomics*, **44**, 455–459.
- Faye, H. and Falzon, P. (2009). Strategies of performance self-monitoring in automotive production. *Applied Ergonomics*, **40**, 915–921.
- Fok, L.Y., Hartman, S.J., Patti, A.L. and Razek, J.R. (2000). Human factors affecting the acceptance of total quality management. *International Journal of Quality & Reliability Management*, **17**(7), 714–729.
- Fuller, C.W. (1999). An employee-management consensus approach to continuous improvement in safety management. *Employee Relations*, **21**(4), 405–418.
- Gallwey, T.J. (1998). Evaluation and control of industrial inspection: Part II – the scientific basis for the guide. *International Journal of Industrial Ergonomics*, **22**(1–2), 51–65.
- Garvare, R. (2002). Process management and sustainable development in a quality perspective – implementation and measurement related to small and medium sized enterprises [doctoral thesis] (Luleå, Sweden: Luleå University).
- Gold, J.E., Punnett, L., Cherniack, M. and Wegman, D.H. (2005). Digital vibration threshold testing and ergonomic stressors in automobile manufacturing workers: A cross-sectional assessment. *Ergonomics*, **48**, 66–77.
- Golomski, W.A. (1994). Human aspects of total quality management. In W. Karwowski (ed.), *Organization and Management of Advanced Manufacturing* (Chicago, IL: John Wiley & Sons, Inc.), p. 103.
- Gonzalez, B.A., Adenso-Diaz, B. and Torre, P.G. (2003). Ergonomic performance and quality relationship: An empirical evidence case. *International Journal of Industrial Ergonomics*, **31**(1), 33–40.
- Goomas, D. and Yeow, P.H.P. (2010). Ergonomics improvement in a harsh environment using an audio feedback system. *International Journal of Industrial Ergonomics*, **40**(6), 767–774.
- Goonetilleke, R.S. and Hoffmann, E.R. (2009). Hand-skin temperature and tracking performance. *International Journal of Industrial Ergonomics*, **39**, 590–595.
- Grandjean, E. (1988). *Fitting the Task to the Man*, 4th edn. (London, U.K.: Taylor & Francis Group).
- Griffin, M.J. (2012). Frequency-dependence of psychophysical and physiological responses to hand-transmitted vibration. *Industrial Health*, **50**, 354–369.
- Hackman, J.R. and Oldham, G.R. (1980). *Work Redesign* (Reading, MA: Addison-Wesley).

- Hackman, J.R. and Wageman, R. (1995). Total quality management: Empirical, conceptual, and practical issues. *Administrative Science Quarterly*, **40**, 309–342.
- Haims, M.C. and Carayon P. (1998). Theory and practice for the implementation of “In-house” continuous improvement participatory ergonomic programs. *Applied Ergonomics*, **29**(6), 461–472.
- Haslegrave, C. (2005). Auditory environment and noise assessment. In J.R. Wilson and E.N. Corlett (eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology*, 3rd edn. (London, U.K.: Taylor & Francis Group), p. 705.
- Helander, M.G. (2006). *A Guide to Human Factors and Ergonomics* (Boca Raton, FL: CRC Press).
- Helander, M.G. and Burri, G.J. (1995). Cost effectiveness of ergonomics and quality improvements in electronics manufacturing. *International Journal of Industrial Ergonomics*, **15**, 137–151.
- Hendrick, H.W. (2008). Applying ergonomics to systems: Some documented “lessons learned”. *Applied Ergonomics*, **39**, 418–426.
- Hendrick, H.W. and Kleiner, B.M. (2001). *Macroergonomics: An Introduction to Work Systems Design* (Santa Monica, CA: Human Factors and Ergonomics Society).
- Hendricks, K.B. and Singhal, V.R. (1997). Does implementing an effective TQM program actually improve operating performance? Empirical evidence from firms that have won quality awards. *Management Science*, **43**(9), 1258–1274.
- Howarth, P.A. (2005). Assessment of the visual environment. In J.R. Wilson and E.N. Corlett (eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology*, 3rd edn. (London, U.K.: Taylor & Francis Group), pp. 663–692.
- Imada, A.S. (1991). The rationale and tools of participatory ergonomics. In K. Noro and A.S. Imada (eds.), *Participatory Ergonomics* (London, U.K.: Taylor & Francis Group), pp. 30–49.
- Imai, M. (1986). *Kaizen: The Key to Japan's Competitive Success* (New York: Random House).
- International Ergonomics Association (IEA). (2014). Definition and domains of ergonomics. Available from: <http://www.iea.cc/whats/index.html> (accessed 12.03.2014).
- ISO 9000. (2008). ISO 9001:2008 Quality management systems requirements. Available from: [http://www.iso.org/iso/iso\\_9000](http://www.iso.org/iso/iso_9000) (accessed 04.03.2014).
- Jorgensen, M., Davis, K., Kotowski, S., Aedla, P. and Dunning, K. (2005). Characteristics of job rotation in the Midwest US manufacturing sector. *Ergonomics*, **48**(15), 1721–1733.
- Juran, J.M. (1974). *Quality Control Handbook* (New York: McGraw-Hill).
- Juslén, H.T. and Tenner, A.D. (2005). Mechanisms involved in enhancing human performance by changing the lighting in the industrial workplace. *International Journal of Industrial Ergonomics*, **35**, 843–855.
- Karia, N. and Asaari, M.H.A.H. (2006). The effects of total quality management practices on employees' work-related attitudes. *TQM Magazine*, **18**(1), 30–43.
- Karlton, J., Axelsson, J. and Eklund, J. (1998). Working conditions and effects of ISO 9000 in six furniture making companies: Implementation and processes. *Applied Ergonomics*, **29**(4), 225–232.
- Karlton, J. and Eklund, J. (1997). Woodworking. In D. Brune, G. Gerhardsson, G.W. Crockford and D. Norbäck (eds.), *The Workplace. Vol. 2: Major Industries and Occupation* (Oslo, Sweden: Scandinavian Science Publisher).
- Kelly, H. and Drury, C. (1999). Socio-technical reasons for the failure of statistical process control. In J. Axelsson, B. Bergman and J. Eklund (eds.), *The International Conference on TQM and Human Factors – Towards Successful Integration*, vol. 1 (Linköping, Sweden: CMTO), pp. 122–127.
- Kim, W.G., Leong, J.K. and Lee, Y.-K. (2005). Effect of service orientation on job satisfaction, organizational commitment, and intention of leaving in a casual dining chain restaurant. *International Journal of Hospitality Management*, **24**(2), 171–193.
- Kirwan, B. and Ainsworth, L.K. (2003). *A Guide to Task Analysis: The Task Analysis Working Group* (London, U.K.: Taylor & Francis Group).
- Klein, J.A. (1984). Why supervisors resist employee involvement. *Harvard Business Review*, **5**, 87–95.
- Klein, J.A. (1989). The human costs of manufacturing reform. *Harvard Business Review*, **67**, 60–66.
- Kogi, K. (2006). Participatory methods effective for ergonomic workplace improvement. *Applied Ergonomics*, **37**, 547–554.
- Kogi, K. (2008). Facilitating participatory steps for planning and implementing low-cost improvements in small workplaces. *Applied Ergonomics*, **39**, 475–481.
- Kondo, Y. (1995). Are creativity and standardization mutually exclusive? *Human System Management*, **14**(4), 309–312.
- Lammermeyr, H.U. (1990). Human relations: The key to quality. *Quality Forum*, **16**, 185–187.
- Lee, K.S. (2005). Ergonomics in total quality management: How can we sell ergonomics to management? *Ergonomics*, **48**(5), 547–558.

- Lee, S.M., Rho, B.-H. and Lee, S.-G. (2003). Impact of Malcolm Baldrige National Quality Award criteria on organizational quality performance. *International Journal of Production Research*, **41**(9), 2003–2020.
- Lillrank, P., Shani, A.B. and Lindberg, P. (2001). Continuous improvement: Exploring alternative organizational designs. *Total Quality Management*, **12**(1), 41–55.
- Lin, L., Drury, C.G. and Kim, S.W. (2001). Ergonomics and quality in paced assembly lines. *Human Factors and Ergonomics in Manufacturing*, **11**(4), 377–382.
- Lovén, E.M. and Helander, M.G. (1997). Effect of operator competence on assessment of quality control in manufacturing. *International Journal of Industrial Ergonomics*, **19**(4), 307–316.
- Lundmark, E. and Westelius, A. (2006). Effects of quality management according to ISO 9000: A Swedish study of the transit to ISO 9000:2000. *Total Quality Management and Business Excellence*, **17**(8), 1021–1042.
- McAtamney, L. and Corlett, E.N. (1993). RULA: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, **24**(2), 91–99.
- Meese, G.B., Kok, R., Lewis, M.I. and Wyon, D.P. (1980). Effect of moderate thermal stress on the potential work performance of factory workers – an interim report. *Energy and Building*, **4**(4), 289–294.
- Melin, U. (1999). The process perspective in total quality management and business process reengineering – a critical review. In J. Axelsson, B. Bergman and J. Eklund (eds.), *The International Conference on TQM and Human Factors – Towards Successful Integration*, vol. 1 (Linköping, Sweden: CMTO), pp. 153–160.
- Molleman, E., van Delft, B. and Slomp, J. (2001). The application of an empowerment model. *Human Factors and Ergonomics in Manufacturing*, **11**(4), 339–354.
- Monden, Y. (1994). *Toyota Production System: An Integrated Approach to Just-in-Time*, 2nd edn. (London, U.K.: Chapman & Hall).
- Moore, J.S. and Garg, A. (1998). The effectiveness of participatory ergonomics in the red meat packing industry: Evaluation of a corporation. *International Journal of Industrial Ergonomics*, **21**, 47–58.
- Munck-Ulfsfält, U., Falck, A., Forsberg, A., Dahlin, C. and Eriksson, A. (2003). Corporate ergonomics programme at Volvo Car Corporation. *Applied Ergonomics*, **34**, 17–22.
- Noro, K. (1988). Participatory ergonomics. *Japanese Journal of Ergonomics*, **24**, 5–10.
- Noro, K. and Imada, A. (eds.) (1991). *Participatory Ergonomics* (London, U.K.: Taylor & Francis Group).
- O'Brien, G. (2002). Participation as the key to successful change – a public sector case study. *Leadership & Organization Development Journal*, **23**(8), 442–455.
- Ooi, K.B., Veeri, A., Yin, L.K. and Vellapan, L.S. (2006). Relationships of TQM practices and employees' propensity to remain: An empirical case study. *TQM Magazine*, **18**(5), 528–541.
- Poksinska, B., Dahlgaard, J.J. and Eklund, J. (2003). Implementing ISO 1400 in Sweden: Motives, benefits and comparisons with ISO 9000. *International Journal of Quality & Reliability Management*, **20**(5), 585–606.
- Rapp, C. and Eklund, J. (2002). Sustainable development of improvement activities – the long-term operation of a suggestion scheme in a Swedish company. *Total Quality Management and Business Excellence*, **13**(7), 945–969.
- Rapp, C. and Eklund, J. (2007). Sustainable development of a suggestion system: Factors influencing improvement activities in a confectionary company. *Human Factors and Ergonomics in Manufacturing & Service Industries*, **17**(1), 79–94.
- Redman, T. and Mathews, B.P. (1998). Service quality and human resource management: A review and research agenda. *Personnel Review*, **27**(1), 57–77.
- Rivera, A.J. and Karsh, B.-T. (2008). Human factors and systems engineering approach to patient safety for radiotherapy. *International Journal of Radiation Oncology, Biology, Physics*, **71**, 174–177.
- Sanders, M.S. and McCormick, E.J. (1993). *Human Factors in Engineering and Design* (New York: McGraw-Hill).
- Schneider, J., Kluth, K. and Strasser, H. (2013). Development and evaluation of specific lighting scenarios and ergonomic approaches to optimise visual inspection tasks. *Theoretical Issues in Ergonomics Science*, **14**, 195–211.
- Seong, Y. and Bisantz, A.M. (2008). The impact of cognitive feedback on judgement performance and trust with decision aids. *International Journal of Industrial Ergonomics*, **38**, 608–625.
- Seppanen, O., Fisk, W.J. and Faulkner, D. (2005). Control of temperature for health and productivity in offices. *ASHRAE Transactions*, **III**(2), 680–686.
- Shiba, S., Graham, A. and Walden, D. (1993). *A New American TQM* (Portland, OR: Productivity Press).
- Shih, R.H., Vasarhelyi, E.M., Dubrowski, A. and Carnahan, H. (2001). The effects of latex gloves on the kinetics of grasping. *International Journal of Industrial Ergonomics*, **28**(5), 265–273.

- Singh, H., Thomas, E.J., Mani, S., Sittig, D., Arora, H., Espadas, D., Khan, M.M. and Petersen, L.A. (2009). Timely follow-up of abnormal diagnostic imaging test results in an outpatient setting: Are electronic medical records achieving their potential? *Archives of Internal Medicine*, **169**, 1578–1586.
- Sittig, D.F. and Singh, H. (2009). Eight rights of safe electronic health record use. *Journal of the American Medical Association*, **302**, 1111–1113.
- Smith, N.C. (1990). *Morality and the Market: Consumer Pressure for Corporate Accountability* (London, U.K.: Routledge).
- Smith-Jackson, T.L. and Klein, K.W. (2009). Open-plan offices: Task performance and mental workload. *Journal of Environmental Psychology*, **29**, 279–289.
- Stollery, B. (1992). Organic solvents. In A.R. Smith and D.M. Jones (eds.), *Handbook of Human Performance* (San Diego, CA: Academic Press), pp. 211–236.
- Tague, N.R. (2005). *The Quality Toolbox* (Milwaukee, WI: American Society of Quality [ASQ] Quality Press).
- Tari, J.J. and Sabater, V. (2006). Human aspects in a quality management context and their effects on performance. *International Journal of Human Resource Management*, **17**(3), 484–503.
- Taylor, W., Melloy, B., Dharwada, P., Gramopadhye, A. and Toler, J. (2004). The effects of static multiple sources of noise on the visual search component of human inspection. *International Journal of Industrial Ergonomics*, **34**(3), 195–207.
- Viaene, M.K. (2002). Overview of the neurotoxic effects in solvent-exposed workers. *Archives of Public Health*, **60**(3–4), 217–232.
- Waterson, P. (2015). Socio-technical design of work systems. In J.R. Wilson and S. Sharples (eds.), *Evaluation of Human Work*, 4th edn. (London, U.K.: CRC Press).
- Westgaard, R.H. and Winkel, J. (1997). Ergonomic intervention research for improved musculoskeletal health: A critical review. *International Journal of Industrial Ergonomics*, **20**, 463–500.
- Wilkinson, A. (1998). Empowerment: theory and practice. *Personnel Review*, **27**(1), 40–56.
- Wilson, J.R. (1991). Participation – a framework and a foundation for ergonomics? *Journal of Occupational Psychology*, **64**(1), 67–80.
- Wilson, J.R. (2014a). Systems ergonomics: Looking into the future – editorial for special issue on systems ergonomics/human factors. *Applied Ergonomics*, **45**(1), 3–4.
- Wilson, J.R. (2014b). Fundamentals of systems ergonomics/human factors. *Applied Ergonomics*, **45**(1), 5–13.
- Wilson, J.R., Haines, H. and Morris, W. (2005). Participatory ergonomics. In J.R. Wilson and E.N. Corlett (eds.), *Evaluation of Human Work: A Practical Ergonomics Methodology*, 3rd edn. (London, U.K.: Taylor & Francis Group).
- Wilson, J.R., Neely, A.D. and Chew, T. (1993). Human and production requirements in modern manufacturing: Complementary or contradictory? *Journal of Design and Manufacturing*, **3**, 167–175.
- Womack, P., Jones, D. and Roos, D. (1990). *The Machine That Changed the World* (New York: Rawson Associates).
- Wright Jr., K.P., Hull, J.T. and Czeisler, C.A. (2002). Relationship between alertness, performance, and body temperature in humans. *American Journal of Physiology – Regulatory Integrative and Comparative Physiology*, **283**(6), 1370–1377.
- Yeow, P.H.P. and Sen, R.N. (2003). Quality, productivity, occupational health and safety and cost effectiveness of ergonomic improvements in the test workstations of an electronic factory. *International Journal of Industrial Ergonomics*, **32**(3), 147–163.
- Yeow, P.H.P. and Sen, R.N. (2004). Ergonomics improvements of the visual inspection process in a printed circuit assembly factory. *International Journal of Occupational Safety and Ergonomics*, **10**(4), 369–385.
- Yeow, P.H.P. and Sen, R.N. (2006). Productivity and quality improvements, revenue increment and rejection cost reduction in the manual component insertion lines through the application of ergonomics. *International Journal of Industrial Ergonomics*, **36**, 367–377.
- Zander, K. and Hamm, U. (2010). Consumer preferences for additional ethical attributes of organic food. *Food Quality and Preference*, **21**(5), 495–503.
- Zegers, D. and van den Berg, V. (1988). The effect of lighting quality and intensity on the efficiency of seed potato inspection. *Zeitschrift für Arbeitswissenschaft*, **42**(2), 106–112.
- Zink, K.J. (1999). Safety and quality issues as part of a holistic (i.e., sociotechnological) approach. *International Journal of Occupational Safety and Ergonomics*, **5**(2), 279–290.
- Zuckerman, A. (1996). Adding value to your ISO 9000 process. *Journal for Quality and Participation*, **19**, 20–22.
- Zwetsloot, G.I.J.M., Aaltonen, M., Wybo, J.L., Saari, J., Kines, P. and De Beeck, R.O. (2013). The case for research into the zero accident vision. *Safety Science*, **58**, 41–48.

---

# 37 Standards in Ergonomics Design and Evaluation

*Mike Gray and Tom Stewart*

## CONTENTS

|                                                                                          |     |
|------------------------------------------------------------------------------------------|-----|
| Introduction.....                                                                        | 957 |
| How Standards Are Developed .....                                                        | 958 |
| Purpose of Standards.....                                                                | 958 |
| Use of Language .....                                                                    | 959 |
| How ISO and CEN Develop Standards.....                                                   | 959 |
| Areas of Ergonomics Standardisation.....                                                 | 960 |
| Human–System Interaction – Usability .....                                               | 960 |
| Early History of ISO Usability Standards.....                                            | 960 |
| How Do You Apply Usability Standards in Practice? .....                                  | 961 |
| Using Standards to Support Human-Centred Design.....                                     | 963 |
| Understand and Specify the Context of Use (Including Users, Tasks and Environments) .... | 963 |
| Specify the User Requirements in Sufficient Detail to Drive the Design .....             | 964 |
| Produce Design Solutions Which Meet These Requirements .....                             | 964 |
| Conduct User-Centred Evaluations of These Design Solutions and Modify the Design         |     |
| Taking Account of the Results.....                                                       | 966 |
| Control Rooms .....                                                                      | 968 |
| Control Rooms .....                                                                      | 969 |
| Biomechanics.....                                                                        | 970 |
| Environmental Ergonomics.....                                                            | 970 |
| Applications: Vehicles, Machine Guarding.....                                            | 971 |
| Getting Involved.....                                                                    | 971 |

## INTRODUCTION

Ergonomics standards (and the standards world tends to use the term ergonomics in preference to human factors), particularly European and International Standards, are powerful tools for design and evaluation. Some of the first ergonomics standards were to do with body size information, human force capabilities and peoples' responses to environmental stressors such as noise and heat. They provided basic data for designers. In more recent times, standards have been developed which aim to support the integration of ergonomics into the design process.

Ergonomics standards are not always easy to find and interpret and are often misunderstood. The purpose of this chapter is to explain what standards can and cannot provide, where to get them and how to apply them to ergonomics/human factors (E/HF) design and evaluation.

Before describing the standards in detail and how they impact on practice, we outline the underlying principles of standardization, especially international standardization.



International standards can be used to improve the quality, effectiveness and safety of goods and services; they are also a facilitator of world trade. International ergonomics standards play an important role in many fields from manufacturing and transport to information systems and information technology, as they encompass widely accepted best practices. However, in order to use them appropriately, it is necessary to understand something about where they come from and how they are developed. The next section in this chapter describes how International Standards and European Standards are developed.

## HOW STANDARDS ARE DEVELOPED

This section starts with explaining the different purposes for standards, some peculiarities about the use of language and the process used by the International Organization for Standardization (ISO) and, within Europe, the European Committee for Standardization (Comité Européen de Normalisation – CEN) to produce standards.

### PURPOSE OF STANDARDS

There are two main purposes for standards – interoperability and best practice. Interoperability standards deal with such issues as defining screw threads in order to ensure that nuts fit bolts and standards for encoding images, for example JPEG, which allow pictures to be shared easily. Best practice standards establish authoritative statements of what should be achieved and can be used to support free trade agreements, providing for goods to meet minimum acceptable levels of performance, for example for safety, if they are to be sold in particular markets.

Ergonomics standards generally fall into the second category and enshrine best practice. They aim to ensure that products, systems or services meeting these standards perform to some specified level. For example, there are standards for lifejackets which aim to ensure that they provide an appropriate level of buoyancy to ensure that people who have to abandon vessels can be kept in a position that allows them to breathe (i.e. face up out of the water) even if they were unconscious. In the usability standards area, there are hardware standards which aim to ensure that computer displays provide an appropriate level of visibility to ensure that people who use them will be able to read them in typical office environments. There are also software usability standards which aim to ensure that dialogue screens are designed to match the users and the tasks they are performing and take account of relevant user psychology and cognitive processes (e.g. ISO 14915 Ergonomic requirements for multimedia user interfaces).

The development of a new standard will normally take several years which might reduce its applicability to the leading edge of design; nonetheless, when properly written, this need not inhibit creativity. Indeed, as discussed later, the development of process standards avoids the trap of limiting standards to current technology. Another approach – which not only does not inhibit creativity but actually enhances it – involves the development of user-based test methods (as exemplified in ISO 9241-304:2008). Demonstrating that the requirements of this standard have been followed involves comparing the performance of a user executing a standard task on a reference display (which meets the standard) and the new display. If the performance on the new display is as good as or better than the reference display, the new display passes. Since the task involves reading and recognizing characters on the display, checking whether the display meets the standard is independent of the technology used. The test simply checks the performance of the user viewing the display. Similar tests have also been agreed for keyboards and other input devices. Far from limiting creativity, these test standards really encourage designers to develop even more innovative technical solutions, which work well for their users.



One of the benefits of E/HF standards is to provide a framework for different parties to share a common understanding when specifying quality in design, procurement and use of products, systems and services. E/HF standards allow

- Customers to set appropriate procurement requirements and to evaluate competing suppliers' offerings
- Suppliers to check their products during design and manufacturing and provide a basis for making claims about the quality of their products
- Regulators to assess quality and provide a basis for testing products against national and international requirements

One of the most significant benefits of European and International standardization is that it places E/HF issues squarely on the agenda. Standards are serious business, and whereas many organizations may choose to pay little regard to research findings, few organizations can afford to ignore standards. Indeed in Europe, and increasingly in other parts of the world, compliance with relevant standards is a key way of demonstrating that legal requirements have been met. Within the European Union (EU), for example the machinery directive has been a major piece of legislation which adopted the approach of developing standards to support the 'Essential Health and Safety Requirements'. This directive is aimed at ensuring goods traded throughout the EU meet certain minimum health and safety standards. This directive makes particular reference to the importance of ergonomics in its requirements, and it is directly as a consequence of this that a number of machinery safety standards were developed to tackle E/HF issues. These include standards on ergonomics design principles, access dimensions, control actuators and displays, force application, hot surfaces, etc.

## USE OF LANGUAGE

It is important to understand that although International Standards are written in English, French or Russian (the three official languages of ISO), there are extensive drafting rules which mean that certain terms have quite specific meanings. For example, in the English versions of the standards, the use of the word 'shall' means that something is necessary in order to meet the requirements in the standard. Meeting all the 'shalls' in the standard is necessary in order to be able to claim that something 'conforms' to the standard.

In ergonomics, there are few circumstances where there is only one solution and therefore where a 'shall' is required. When designers ask E/HF experts for advice, the answer usually starts with 'it depends...'. The reason for this is that for many decisions within E/HF, the 'correct answer' depends on the context. So, for example, an apparently simple question such as 'what is the maximum number of items that can be presented as a menu?' does not have a simple answer. If the menu is a simple drop-down list which the user will scroll down using a mouse or arrow key, then according to ISO 9241-14:2000 Menu Dialogues, that number is generally 7 plus or minus 2 corresponding to short-term memory. However, menus which can be refined by entering some initial text (e.g. stations for a train timetable) could be as long as the number of entries.

In practice, therefore, ergonomics standards generally provide 'guidance' and use the word 'should' (in English versions) to reflect that this is a 'recommendation' rather than a 'requirement'.

## HOW ISO AND CEN DEVELOP STANDARDS

ISO is the world's largest developer and publisher of International Standards. Its members are the National Standards Bodies (NSBs) of 162 countries. It is supported by a Central Secretariat based in Geneva, Switzerland. The principal deliverable of ISO is the International Standard (IS).

CEN is a major provider of European Standards and technical specifications. Its members are the NSBs from 33 countries and its principal deliverable is the European Standard (EN).

These standards have a unique status since they also are national standards in each of its 33 member countries.

The Vienna Agreement – signed by CEN in 1991 with ISO – aims to ensure that the same text is published as both an ISO Standard and a European Standard.

In both bodies, standards development work is conducted by technical committees (TCs). ISO TC159 and CEN TC122 are responsible for ergonomics standards. In practice, the technical work takes place in Working Groups (WGs) of experts, nominated by national standards committees but expected to act as independent experts. Most NSBs set up ‘mirror committees’ to coordinate their input to the international work.

The standards progress through various stages from initial new work item proposals to finally approved and published standards, and this can take between 3 and 5 years to reach sufficient consensus for a final text to be agreed. Voting on draft standards is co-ordinated by the NSBs, and sufficient agreement has to be gained before an International Standard can be published. The ISO website provides more information on this process ([www.iso.org](http://www.iso.org)).

Standards are reviewed at regular intervals both to see if they are still needed and whether their content should be updated.

## **AREAS OF ERGONOMICS STANDARDISATION**

ISO and CEN standards are the responsibility of more than 100 TCs dealing with a wide range of standardisation areas ranging from ISO TC1 Screw Threads to ISO TC/PC 283 Occupational Health and Safety Management Systems. The TCs responsible for Ergonomics (ISO TC159 and CEN TC122) cover a wide area of ergonomics from usability and the human-centred design process to machinery guarding and manual handling. Each of the TCs’ WGs is responsible for developing individual standards in a particular area, for example anthropometry or control centres. The following section describes the standards in five main areas and discusses how standards can be used in ergonomics design and evaluation.

### **HUMAN–SYSTEM INTERACTION – USABILITY**

ISO and CEN usability standards have become increasingly important in the design of information systems and technology. They are specified in many procurement contracts, and most major software developers rely on them to demonstrate the quality of their products. Their history stretches back over more than 30 years.

### **EARLY HISTORY OF ISO USABILITY STANDARDS**

Usability standards have been under development by ISO since 1983 when a new work item with the title ‘Visual Information Processing’ was approved. This work item became the responsibility of ISO TC159 who set up a sub-committee (SC4) ‘Ergonomics of Human-System Interaction’ to develop standards in the area. After several years work, this resulted in a 17-part standard, ISO 9241, titled ‘Ergonomic requirements for office work with visual display terminals’. The standard was also adopted as EN ISO 9241. The focus on office work reflected concerns at the time about the impact of VDU use on people’s health, but many of the parts of ISO 9241 dealt with broader usability and software interface issues. It has been an influential standard across the world and is referenced in the regulations of many countries, including the United Kingdom’s Health and Safety (Display Screen Equipment) Regulations 1992.

When it came to revising this standard, the experts involved in developing it wanted both to retain the ISO 9241 ‘brand’, which had become a recognized benchmark and also to broaden its scope. A new title *The Ergonomics of Human-System Interaction* reflected this ambition, and the multi-part structure allowed the committee to integrate other usability standards.

The structure reflected the numbering of the original ISO 9241 standard, for example displays were originally Part 3 and became the 300 series. In each section, the ‘hundred’ is an introduction to the section, for example Part 100 gives an introduction to the software ergonomics parts. The number of parts which will be developed in each section varies depending on the complexity of the area and the need for specific standards.

Only three part numbers have been retained (following revision) from the original ISO 9241 structure, as these address issues which apply across all other parts of ISO 9241:

Part 1: Introduction

Part 2: Job design

Part 11: Hardware and software usability

An additional part which also applies across all of ISO 9241 is

Part 20: Accessibility and human–system interaction

The remaining parts are structured in ‘hundreds’ as follows:

100 series: Software ergonomics

200 series: Human–system interaction processes

300 series: Displays and display related hardware

400 series: Physical input devices – ergonomics principles

500 series: Workplace ergonomics

600 series: Environment ergonomics

700 series: Application domains – control rooms (incorporating ISO 11064)

900 series: Tactile and haptic interactions

## HOW DO YOU APPLY USABILITY STANDARDS IN PRACTICE?

During the development of ISO 9241, it became clear that user interface technology was developing so quickly that the traditional approach to standards could not keep up in all areas. Two solutions were adopted for this problem.

First, standards concerned with hardware including displays and input devices contained what were known as ‘user performance test methods’ as well as conventional hardware test methods. These user performance test methods allowed a new device to be used under controlled conditions by a defined group of participants performing specific tasks. The performance of the participants was then compared with their performance on a test device, which met the standard. If the participants’ performance was as good or better on the test device, then this device was deemed to pass. This approach means that any technology can be tested if it is designed to allow users to perform suitable tasks, for example read text on a screen or enter characters into a system.

Second, it was realized that, in many cases, the best way to ensure that a design was usable was to follow a human-centred design process. Human-centred design is an approach to developing and acquiring software, which improves system effectiveness and the efficiency and satisfaction of users (usability). Following this systematic framework helps reduce the risk that new software is unusable or fails to work properly for its intended users. Across all industries and governments, major software development projects are notorious for being delivered late and over budget. Worse still, they often fail to deliver the service promised. In hindsight, the problems are often due to a failure to engage fully with all stakeholders, especially those intended to operate and use the systems.

Furthermore, excessive ambition which overlooks the reality of existing working practices, design and development methods which are rigid, inflexible and unable to adapt to emerging business needs and changes in the environment are also reasons for failure. In 1999, ISO TC159 SC4 published a new standard ISO 13407 on human-centered design for interactive systems. The standard provided a systematic framework, which directly addressed these and other problems in delivering usable software, products and services. It was aimed at project managers and described how they can apply human-centred design processes to procure or develop technical software which will deliver value to the users and the business. This standard presented four principles of human-centred design, including the iteration of design solutions and the active involvement of users. It also described key activities including the need to understand and specify the context of use, produce designs and prototypes of potential solutions and evaluate these solutions against the human-centred criteria. The term human-centred design was used in preference to user-centred design because the process includes all stakeholders, not just those usually considered as users.

Both approaches have been adopted widely, and the new parts of ISO 9241 dealing with displays and input devices both contain user performance test methods. Human-centred design has been recognized as the best way to develop usable systems, and the standard has been endorsed by such bodies as the User Experience Professionals Association (UXPA) and the International Ergonomics Association (IEA).

The human-centred design approach is complementary to existing design methodologies – it does not describe all the processes which are necessary to design and develop a system. It is characterized by the following:

- Systems are designed to take account of all the people who will use them and other stakeholder groups, including those who might be affected (directly or indirectly) by their use. Constructing systems based on an inappropriate or incomplete understanding of user needs is one of the major sources of systems failure.
- Users are involved throughout design and development. This provides knowledge about the context of use, the tasks and how users are likely to work with the future system. User involvement should be active, whether by participating in design, acting as a source of relevant data or by evaluating solutions.
- The design is driven and refined by user-centred evaluation. Early and continuous feedback from users minimizes the risk that the system does not meet user and organizational needs (including those requirements that are hidden or difficult to specify explicitly). Such evaluation ensures preliminary design solutions are tested against ‘real-world’ scenarios, with the results being fed back into progressively refined solutions. User-centred evaluation should also take place as part of final acceptance of the product to confirm that requirements have been met. Feedback from users during operational use identifies long-term issues and provides input to future design.
- The process is iterative. For complex systems, it is impossible to specify completely and accurately every aspect of the interaction at the beginning of development. Initial design solutions rarely satisfy all the user needs. Many user and stakeholder needs only emerge during development, as the designers refine their understanding of the users and their tasks and, as users, provide feedback on potential solutions.
- The design addresses the whole user experience. This involves considering, where appropriate, organizational impacts, user documentation, online help, support and maintenance (including help desks and customer contact points), training and long-term use.
- The design team reflects a range of skills and perspectives. The team does not have to be large, but should be sufficiently diverse to understand the constraints and realities of the various disciplines involved. This can have other benefits too. For example, technical experts can become more sensitized to user issues, and the users themselves can become more aware of technical constraints.

## USING STANDARDS TO SUPPORT HUMAN-CENTRED DESIGN

Although the human-centred design standard ISO 13407 was developed in parallel with ISO 9241, many parts of ISO 9241 were developed with an implicit human-centred design process in mind. These parts can also be used to support specific design activities whether or not a full human-centred design process is being followed. Therefore, the next section describes the human-centred design process, originally specified in ISO 13407, and explains how the various parts of ISO 9241 can be used to support this process.

As part of the overall restructuring of ISO 9241 mentioned earlier, the opportunity was taken when ISO 13407 was undergoing its systematic review (a key part of the ISO process to ensure continued relevance) to incorporate it within the new ISO 9241 structure as ISO 9241-210:2010 human-centred design for interactive systems. The revised standard remained largely unchanged, with one major exception. The four key human-centred design activities were no longer just recommendations; they became ‘requirements’. This meant that people could now claim conformance, that is they followed and adhered to the ISO 9241-210 human-centred design process. To do so, they have to ensure that they

- Understand and specify the context of use (including users, tasks and environments)
- Specify the user requirements in sufficient detail to drive the design
- Produce design solutions which meet these requirements
- Conduct user-centred evaluations of these design solutions and modify the design taking account of the results

Note that the requirement is to provide evidence to show that they have done these successfully. There is no requirement to produce lots of documents to conform to the standard. The onus is on the organization claiming conformance to provide whatever level of detailed evidence the recipient requires.

Standards development is an ongoing process, and standards can change radically during the various stages. The tables therefore only reference standards which have been published up to mid-2012.

### Understand and Specify the Context of Use (Including Users, Tasks and Environments)

The characteristics of the users, their tasks and the organizational, technical and physical environment define the context in which the system is used (the context of use).

The context of use description includes details of the following:

- Users and other stakeholder groups
- User profile in terms of relevant knowledge, skill, experience, training, physical attributes (e.g. any disabilities), preferences and capabilities of the users
- Environment profile in which the system will operate: the technical environment, including the hardware, software and materials, the relevant characteristics of the physical environment including thermal conditions, lighting, spatial layout and furniture
- User task profile including the frequency and duration of tasks and interdependencies

The user, environment and task profiles should be described in sufficient detail to support the requirements, design and evaluation activities. They are working documents that start as outlines and are reviewed, maintained, extended and updated during the design and development process.

The next step is to turn the profiles into concrete, detailed usage ‘scenarios’ and ‘personas’. Scenarios are realistic stories that describe the use of the software from the user’s perspective and are combinations of individual tasks in the right sequence to achieve an overall goal. Personas are rich descriptions of target users.

Table 37.1 shows standards which support understanding and specifying the context of use.

**TABLE 37.1**  
**Understand and Specify the Context of Use**

| Relevant Parts of ISO 9241                                                                  | Brief Description                                                                                                                                                                                                                                             |
|---------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO 9241-210: 2010 Human-centred design for interactive systems (supersedes ISO 13407 1999) | Guidance on and requirements for the human-centred design process, aimed at HCD managers                                                                                                                                                                      |
| ISO 9241-11:1998 Guidance on usability                                                      | Defines usability as ‘Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use’ and provides guidance on how to address usability in design projects |

**Specify the User Requirements in Sufficient Detail to Drive the Design**

Identifying user needs and specifying the functional and other requirements for the system are major software development activities. When following the human-centred design approach, this activity is extended to create an explicit specification of user requirements in relation to the context of use and the objectives of the system.

The user requirements specification should be

- Stated in terms that permit subsequent testing
- Verified by the relevant stakeholders
- Internally consistent
- Updated as necessary, during the life of the project

This specification is then used to drive the design and provide agreed quality measures used to evaluate the software.

Table 37.2 shows standards which support specifying the user requirements.

**Produce Design Solutions Which Meet These Requirements**

Within the ISO 9241 series of standards, there are several standards which provide detailed guidance, but the overall process involves six main steps in designing usable software.

1. Structure solutions around key tasks and workflows based on the context of use and user requirements specification. A key issue concerns establishing an appropriate ‘allocation of function’ – deciding which system tasks are to be automated and which should be under user control.
2. Design the interaction, interface and navigation from the user’s perspective and keep them consistent. Be aware of other software that the user is likely to be using and be cautious about introducing different styles of interface. Table 37.3 lists standards relevant to designing user interfaces and navigation.

**TABLE 37.2**  
**Specify the User Requirements in Sufficient Detail to Drive the Design**

|                                                                                                                                    |                                                                                                                                                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO 9241-20:2009 Accessibility guidelines for information/communication technology (ICT) equipment and services                    | High-level guidance on the design and selection of equipment and services for people with a wide range of sensory, physical and cognitive abilities, including those who are temporarily disabled and the elderly. |
| ISO 9241-2:1992 Guidance on task requirements deals with the design of tasks and jobs involving work with visual display terminals | Guidance on identifying and specifying task requirements and how they can be incorporated into the system design and implementation process.                                                                       |



**TABLE 37.3****Produce and Evaluate (Software) Design Solutions**

|                                                                               |                                                                                                                                                                                                                                                                                                                                                                                      |
|-------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO/TR 9241-100:2010 Introduction to standards related to software ergonomics | Introduction to the '100 series' of ISO 9241 software parts.                                                                                                                                                                                                                                                                                                                         |
| ISO 9241-110:2006 Dialogue principles (supersedes ISO 9241 10:1996)           | Sets out seven user interface principles – suitability for the task; self-descriptiveness; controllability; conforms to user expectations (i.e. consistent); error tolerant and forgiving; suitable for individualization and customizing; and support learning.                                                                                                                     |
| ISO 9241-129:2010 Guidance on software individualization                      | Guidance on where individualization might be appropriate or inappropriate and how to apply individualization.                                                                                                                                                                                                                                                                        |
| ISO 9241-14:2000 Menu dialogues                                               | Recommends best practice for designing menus (pop-up, pull-down and text-based menus). Topics include menu structure, navigation, option selection and menu presentation. Includes 10-page checklist for determining compliance.                                                                                                                                                     |
| ISO 9241-15:1998 Command dialogues                                            | Recommends best practice for designing command languages covering command language structure and syntax, command representations, input and output considerations, and feedback and help.                                                                                                                                                                                            |
| ISO 9241-16:1999 Direct manipulation dialogues                                | Recommends best practice for designing direct manipulation dialogues and includes the manipulation of objects, and the design of metaphors, objects and attributes.                                                                                                                                                                                                                  |
| ISO 9241-143:2012 Forms                                                       | Provides requirements and recommendations for the design and evaluation of screen-based forms. Includes guidance on the selection and design of interface elements relevant to forms.                                                                                                                                                                                                |
| ISO 9241-151:2008 Guidance on World Wide Web user interfaces                  | Sets out principles for designing usable websites – these cover high-level design decisions and design strategy, content design, navigation and content presentation.                                                                                                                                                                                                                |
| ISO 9241-154: Interactive voice response (IVR) applications                   | Provides guidance on, and requirements for, the user interface design of interactive voice response (IVR) applications. Covers systems that employ touchtone input as well as those using automated speech recognition (ASR) as the input mechanism.                                                                                                                                 |
| ISO 9241-171:2008 Guidance on software accessibility                          | Provides guidance on the design of software to achieve as high a level of accessibility as possible. Replaces the earlier Technical Specification ISO TS 16071:2003 and follows the same definition of accessibility – 'usability of a product, service, environment or facility by people with the widest range of capabilities'. Applies to all software, not just web interfaces. |

3. Define the navigation and keep it consistent. If different teams are developing separate parts of the software, maintain consistency by establishing an agreed style guide at the beginning of the development.
4. Follow interface design best practice. Various parts of ISO 9241 offer guidance and are based on internationally agreed best practice for achieving usable hardware and software. Tables 37.4 through 37.6 address input devices, displays and workplaces.
5. Produce sketches and mock-ups early to test assumptions. Using scenarios and simple sketches or mock-ups enables the designers to communicate the proposed design to users and other stakeholders to obtain feedback before design decisions are finalized.
6. Keep testing emerging solutions with users until the quality criteria are met. Human-centred evaluation can take place throughout the design process from initial concepts to final (signed-off) designs.

**TABLE 37.4****Parts of ISO 9241 Relevant to Designing or Selecting Displays**

| <b>Relevant Part of ISO 9241</b>                                                      | <b>Brief Description</b>                                                                                                                                                                                                                                                                                                                              |
|---------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO 9241-300:2008 Introduction to electronic visual display requirements              | A very short (four pages) introduction to the ISO 9241-300 series which explains what the other parts contain.                                                                                                                                                                                                                                        |
| ISO 9241-302:2008 Terminology for electronic visual displays                          | Definitions, terms and equations that are used throughout ISO 9241:300 series.                                                                                                                                                                                                                                                                        |
| ISO 9241-303:2011 Requirements for electronic visual displays                         | Sets general image quality requirements for electronic visual displays. The requirements are intended to apply to any kind of display technology.                                                                                                                                                                                                     |
| ISO 9241-304:2008 User performance test methods for electronic visual displays        | Describes methods which involve testing how people perform when using the display. The method can be used with any display technology.                                                                                                                                                                                                                |
| ISO 9241-305:2008 Optical laboratory test methods for electronic visual displays      | Defines optical test methods and expert observation techniques for evaluating a visual display against the requirements in ISO 9241-303.                                                                                                                                                                                                              |
| ISO 9241-306:2008 Field assessment methods for electronic visual displays             | Provides guidance on how to evaluate visual displays in real-life workplaces.                                                                                                                                                                                                                                                                         |
| ISO 9241-307:2008 Analysis and compliance test methods for electronic visual displays | Supports ISO 9241-305 with very detailed instructions on assessing whether a display meets the ergonomics requirements set out in Part 303.                                                                                                                                                                                                           |
| ISO/TR 9241-308:2008 Surface-conduction electron-emitter displays (SED)               | Technical report on a new 'eco-friendly' display technology called 'Surface-Conduction Electron-Emitter Displays' (SED).                                                                                                                                                                                                                              |
| ISO/TR 9241-309:2008 Organic light-emitting diode (OLED) displays                     | Technical report on another new display technology called 'Organic Light Emitting Diode Displays' (OLED) which are better for fast-moving images than LCDs.                                                                                                                                                                                           |
| ISO/TR 9241-310:2010 Visibility, aesthetics and ergonomics of pixel defects           | Explains why some pixel defects are inevitable in low-cost displays and gives guidance on the specification of pixel defects, visibility thresholds and aesthetic requirements.                                                                                                                                                                       |
| ISO/TR 9241-331:2012 Optical characteristics of autostereoscopic displays             | Addresses the optical properties of autostereoscopic displays (ASDs) in order to reduce visual fatigue. Provides terminology, performance characteristics and optical measurement methods for ASDs. Applicable to spatially interlaced autostereoscopic displays (two-view, multi-view and integral displays) of the transmissive and emissive types. |
| ISO 9241-12:1999 Presentation of information                                          | Guidance on ways of representing complex information using alphanumeric and graphical/symbolic codes, screen layout and design as well as the use of windows.                                                                                                                                                                                         |

### **Conduct User-Centred Evaluations of These Design Solutions and Modify the Design Taking Account of the Results**

User-centred evaluation (evaluation based on the user's perspective) is at the heart of human-centred design. It occurs throughout the design life cycle from initial testing of concepts to final testing to confirm that the software meets the quality criteria.

There are two main approaches to user-centred evaluation – inspection-based evaluation and user-based testing.

Inspection-based evaluation is ideally performed by usability experts, with broad experience of problems encountered by users, working with domain experts who understand the software and the problem domain. Inspection can be based on task scenarios or on guidelines and standards, for example the Web Content Accessibility Guidelines (WCAG).

During the development process or where there are no users already familiar with the software, inspection using scenarios can be used. This involves the assessor putting themselves in the position of the user working through the scenarios agreed in the requirements stage. This can be done with working prototypes or with partial prototypes or even non-functioning mock-ups of the software.

**TABLE 37.5****Parts of ISO 9241 Relevant to Designing or Selecting Keyboards and Other Input Devices**

| Relevant Part of ISO 9241                                                                                                   | Brief Description                                                                                                                                                                                                                                 |
|-----------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO 9241-4:1998 Keyboard requirements (some clauses in this standard have been superseded by ISO 9241-400 and ISO 9241-410) | Specifies the ergonomics design characteristics of an alphanumeric keyboard, which may be used comfortably, safely and efficiently to perform office tasks.                                                                                       |
| ISO 9241-400:2007 Principles and requirements for physical input devices                                                    | Sets out the general ergonomics principles and requirements, which should be taken into account when designing or selecting physical input devices.                                                                                               |
| ISO 9241-410:2008 Design criteria for physical input devices (supersedes ISO 9241-9: 1998)                                  | Describes ergonomics characteristics for input devices, including keyboards, mice, pucks, joysticks, trackballs, touchpads, tablets, styli and touch-sensitive screens.                                                                           |
| ISO/TS 9241-411:2012 Evaluation methods for the design of physical input devices                                            | Describes ergonomics evaluation and test methods for physical input devices including keyboards, mice, pucks, joysticks, trackballs, touchpads, tablets and overlays, styli and light pens and touch-sensitive screens.                           |
| ISO 9241-420:2011 Selection of physical input device                                                                        | Provides ergonomics guidance for selecting input devices for interactive systems taking account of the limitations and capabilities of users, the tasks and the context of use when selecting input devices.                                      |
| ISO 9241-910:2011 Framework for tactile and haptic interaction                                                              | Defines terms, describes structures and models, and also serves as an introduction to the other parts of the ISO 9241 '900' subseries. It provides guidance on how tactile/haptic interaction can be applied to a variety of user tasks.          |
| ISO 9241-920:2009 Guidance on tactile and haptic interactions                                                               | Gives recommendations on the design and evaluation of hardware, software and combinations of hardware and software interactions. It does not provide recommendations specific to Braille, but can apply to interactions that make use of Braille. |

**TABLE 37.6****Parts of ISO 9241 Relevant to Designing or Workplaces for Display Screen Users**

| Relevant Part of ISO 9241                                     | Brief Description                                                                                                                                         |
|---------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO 9241-5:1999 Workstation layout and postural requirements. | Specifies the ergonomics requirements for a visual display terminal workplace which will allow the user to adopt a comfortable and efficient posture.     |
| ISO 9241-6:2000 Guidance on the work environment.             | Specifies the ergonomics requirements for a VDT working environment which will provide the user with comfortable, safe and productive working conditions. |

Inspection-based evaluation can also be conducted using guidelines and standards. Software accessibility, the degree to which software can be used by people with disabilities, is one area where it can be necessary to check compliance with published external standards. In particular, the WCAG, published by the World Wide Web Consortium (W3C), are widely accepted accessibility standards for web-based applications.

User testing can be undertaken at any stage of the design and development cycle. At an early stage, users can be presented with sketches or mock-ups of design concepts and asked to evaluate them by working through how they would carry out key tasks. This provides much more information than simply showing or demonstrating them to users. For instance, scenarios could be used to provide a useful framework for structuring this walk through of the software. Later in development, user testing is used to confirm that the software meets the agreed quality criteria established earlier.

User testing is best conducted in a usability laboratory which allows user behaviour including detailed mouse and keyboard use to be observed and recorded for subsequent analysis. This can be supplemented with eye-tracking technology, which shows exactly where on the screen the users have been looking during the tasks. It can also be conducted more simply at the user's normal workstation by observing and video recording their behaviour as they work through the tasks.

In addition to direct observations or measurements of user behaviour (e.g. task completion times), user testing can involve users vocalizing their thought process while they perform the tasks. Further feedback can be obtained through questionnaires or interviews and checklists (before, during or after the tasks) (see Chapter 13 for further discussion of HCI evaluation methods).

### Control Rooms

When ISO 9241 was first being developed in 1983, most control rooms contained custom-built workstations and control equipment specially designed for that specific industry. For example, railway signal boxes still used mechanical levers and indicator lights, and power control room had massive display/control panels with meters, buttons and lights.

The potential for visual display terminals (VDTs) to provide some of this functionality in such environments was already being recognized. However, it was clear that the tasks and priorities of a control room operator were quite different from the typical office tasks and workers that were the focus of ISO 9241. Although productivity was important in both environments, safety quite rightly dominated control room design. The well-being of control room staff inevitably came second.

It was therefore decided that although many of the requirements and much of the guidance in ISO 9241 were relevant to control room design, especially where the use of VDTs was concerned, there was a need for a new standard to address the specific needs of control rooms.

This multipart standard (ISO 11064) adopts a user-centred approach but recognizes that control rooms form part of a larger more complex system. Table 37.7 shows the seven parts of ISO 11064.

**TABLE 37.7**

#### Parts of ISO 11064 Relevant to Designing and Evaluating Control Rooms

| Relevant Part of ISO 11064                                        | Brief Description                                                                                                                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ISO 11064-1:2000 Principles for the design of control centres     | High-level ergonomics principles for designing control centres which house control suites to ensure safe and effective working conditions.                                                                                                                                                                                                                         |
| ISO 11064-2:2000 Principles for the arrangement of control suites | High-level ergonomics principles for laying out control suites within control centres.                                                                                                                                                                                                                                                                             |
| ISO 11064-3:2000 Control room layout                              | Specific space planning and design guidance for laying out control rooms within control suites.                                                                                                                                                                                                                                                                    |
| ISO 11064-4:2013 Layout and dimensions of workstations            | Covers workstation design with particular emphasis on layout and dimensions. Relevant to workstations found in applications such as transportation control, process control and security installation.                                                                                                                                                             |
| ISO 11064-5:2008 Displays and controls                            | Requirements and recommendations for displays, controls and their interaction, in the design of control-centre hardware and software.                                                                                                                                                                                                                              |
| ISO 11064-6:2005 Environmental requirements for control centres   | Provides environmental requirements and recommendations for the ergonomic design, upgrading or refurbishment of control rooms and other functional areas within the control suite. The following aspects are covered: thermal environment (temperate regions), air quality, lighting environment, acoustic environment, vibration, aesthetics and interior design. |
| ISO 11064-7:2006 Principles for the evaluation of control centres | Provides requirements, recommendations and guidelines on evaluation of the different elements of the control centre, that is control suite, control room, workstations, displays and controls and work environment.                                                                                                                                                |

Over the years, care has been taken to ensure that there are no conflicts between ISO 9241 and ISO 11064. This is particularly important in newer control rooms, which often use 'off-the-shelf' VDTs and even regular office desks. Even in such cases, the control rooms are more likely to have multi-screen workstations, and the system safety aspect of the control room must remain a priority. In addition, even though there is a convergence between offices (many becoming multi-screen) and control rooms (many looking very like offices), few offices operate 24/7 and this not only places more demands on the people, equipment and furniture but also makes it even more important to get the ergonomics right.

When ISO 9241 was being restructured, the 700 series numbers were reserved for specific application areas. One of these was control rooms. However, since the ISO 11064 series was still under development, it was agreed that the standard should be completed before any re-numbering should be considered. Now that ISO 11064 is complete, the possibility of renumbering to include it into the ISO 9241 series is being reviewed. ISO 11064 is now a recognized 'brand' in many industries and changing the number might only serve to confuse its users. Furthermore, following a user-centred design process (as described in ISO 9241-210) suggests that the interests of the users of the standard should be a higher priority than the desire of the standards makers for a logical structure. Indeed, in the fast-moving information technology industry, achieving a logical structure is never going to be easy or may be even possible. Technology has a habit of making simplistic distinctions irrelevant, for example it used to make sense to distinguish input and output devices – so where does that leave a tablet? The final location of ISO 11064 is still to be resolved, but in the meantime, the standard continues to provide much useful ergonomics design guidance for control room designers.

## CONTROL ROOMS

The development of standards for anthropometry was one of the earliest activities of the TC for ergonomics. Standardized definitions of body measurements are important so that results from anthropometric surveys are comparable and users of the data can have confidence in their results. Because surveys are costly to perform and measurement techniques can need skilled people to perform them, standards have also been drawn up which describe the essentials for a reliable survey.

ISO 7250 'Basic human body dimensions for technological design' has several parts: Part 1 provides definitions of dimensions and describes the body landmarks to be used and the methods for taking the measurements and Part 2 provides summary statistical data from national surveys for the dimensions defined in Part 1: it is intended to be updated regularly as suitable survey data are made available.

ISO 15535 'General requirements for establishing anthropometric databases' describes those aspects which should ensure that an anthropometric survey provides reliable and useable data. A standard on the requirements for using 3-D scanning methods to obtain anthropometric data which is compatible with other survey methods (ISO 20685) has also been produced.

There are also standards which give information on how to use anthropometric data. These include sizing of garments, design of test dummies for vehicles tests and the specification of tests for reach zones and sight lines. ISO 15536 standards describe how templates and computer manikins should be specified and used so that they can represent human body dimensions.

Methods for carrying out tests on pieces of equipment using test panels of users are specified in ISO 15537.

Safety requirements can be addressed by several of these standards: ISO 14738 'Safety of machinery – Anthropometric requirements for the design of workstations at machinery' was written primarily to address essential requirements of the European Machinery Directive. It specifies how to take account of people's body dimensions when designing sitting, standing and sit/stand work areas.

## BIOMECHANICS

Standards on manual handling and biomechanics have been developed in parallel within CEN and ISO but with somewhat different applications in mind. The European standard EN1005 'Safety of Machinery – Human Physical Performance' provides four parts covering the requirements on handling parts of a machine to meet the needs of the machinery directive. It requires machinery designers to adopt a three-stage approach: to avoid manual handling activities wherever possible, to utilize technical aids and to further reduce the inherent level of risk by optimizing handling activities. ISO 11228 'Ergonomics – manual handling' uses a very similar approach to provide standards on Lifting and carrying (Part 1), Pushing and pulling (Part 2) and Handling of low loads at high frequency (Part 3) – these standards are applicable to all types of handling.

A further standard ISO 11226 'Ergonomics – Evaluation of static working postures' provides methods for assessing the postures that people use while at work.

The WG responsible for these standards has also produced a technical report (TR) ISO TR 12296 Ergonomics – Manual handling of people in the healthcare sector. This provides information on using the approaches of risk estimation and risk evaluation of manual handling in a health care context.

The development of standards for biomechanics has helped to establish internationally agreed methodologies for assessing tasks where people are expected to exert forces. The intention has been to provide simple or screening methods which can be used for straightforward assessments but also to provide comprehensive, more analytical, methods where more complex tasks need to be evaluated.

## ENVIRONMENTAL ERGONOMICS

Standards in this category are often developed by specialists in lighting, noise, heating and ventilation, etc. However, ergonomics specialists have developed standards which address the human performance aspects of environmental conditions.

The ergonomics of the thermal environment is particularly well developed in its approach to producing standards. The ISO WG responsible for this topic has produced standards on several indices of heat stress and discomfort (e.g. ISO 7933 'Analytical heat stress', ISO 7243 'WBGT – Index', ISO 9888 'Evaluation of thermal strain by physiological measurements' as well as standards on approaches to measuring thermal environments (ISO 7726), cold environments (ISO 11079) and thermal comfort (EN 7730)).

Methods for assessing human responses to both hot and cold surfaces have been produced and published in ISO 13732. Threshold limit values and a risk assessment procedure for preventing burns when the skin is in contact with a hot solid surface are given in Part 1 of this standard.

Standards for lighting of workplaces provide lighting design criteria for providing sufficient illumination and avoiding glare and other adverse effects on people. Requirements for indoor workplaces are given in EN 12464-1, while outdoor workplaces are covered in EN 12464-2. In addition, the Society for Light and Lighting and the Commission Internationale de L'éclairage (CIE) provide detailed Lighting Codes for particular environments.

Standards dealing with the ergonomics aspects of noise have dealt with the issues of communicating in noisy environments. ISO 7731 deals with danger signals in public and work spaces, while ISO 9921 gives information on assessing the quality of speech communication when warning of danger or conveying information messages, for example in public areas.

However, ergonomics is usually interested in the effect on people of real-life environmental challenges, and these rarely occur in isolation. ISO have set up a WG under the ergonomics TC to tackle the issues involved with combined effects of environmental components. ISO 28802 'Ergonomics of the physical environment – Assessment of environments by means of an environmental survey involving physical measurements of the environment and subjective responses



of people' is intended to allow the assessment of human response to the total environment. This includes thermal, acoustic, lighting and air quality environments.

Although ergonomists will typically try to accommodate a wide range of the population's abilities and limitations, standards makers have also been charged with considering people with special requirement. ISO Guide 71 'Guidelines for standards developers to address the needs of older persons and persons with disabilities' sets out how ISO intends to implement its policies in this area. Ergonomists are well placed to provide approaches for enabling such work. Standards providing ergonomics requirements for accessible design have been produced: for example ISO 28803 on the physical environment and ISO 24500 on auditory signals for accessible design.

### **APPLICATIONS: VEHICLES, MACHINE GUARDING**

Many product standards contain requirements for considering ergonomics aspects during design. These may refer to more general standards on ergonomics or build-in specific requirements in the standards themselves. Vehicle design makes much use of human manikins for testing particular aspects of performance. The design of manikins is intended to represent features of the user population. For example, dummies used for crash testing of vehicles need to behave sufficiently like real people in terms of their size, weight, resilience and dynamics. ISO 15830 specifies the World Side Impact Dummy characteristics. Earth moving machine standards provide requirements for access dimensions (ISO 2860).

Mechanical guards are meant to ensure that people cannot reach dangerous parts of a machine. The placement and design of the guards need to take account of the dimensions of those who might reach furthest or can reach through the smallest gaps. ISO 13857 'Safety of machinery Safety distances to prevent hazard zones being reached by upper and lower limbs' provides tables of data for selecting barrier heights and guard distances which will put hazards out of reach. Information on how far through openings of various sizes and shapes someone could reach is also provided.

Technological developments can require standards to be developed in new areas. For example, as collaborative robots are developed which work directly alongside people, then standards are also being developed to consider how adequate safeguards can be put in place to ensure people are unlikely to be injured. Consideration of people's abilities to sustain impacts on different parts of the body need to be taken into account when providing solutions which will help achieve such protection.

### **GETTING INVOLVED**

Any readers who wish to get involved in the international standardization processes described in this chapter should contact their NSB to find the best way they can contribute. This could be by direct involvement as a national expert on one of the ISO TC159 Ergonomics WGs or by providing comments on documents via the relevant national committee.



# *Section VII*

---

## *Conclusion*



---

# 38 Ergonomics and Human Factors as Reflective Practice

*Sarah Sharples and Peter Buckle*

## CONTENTS

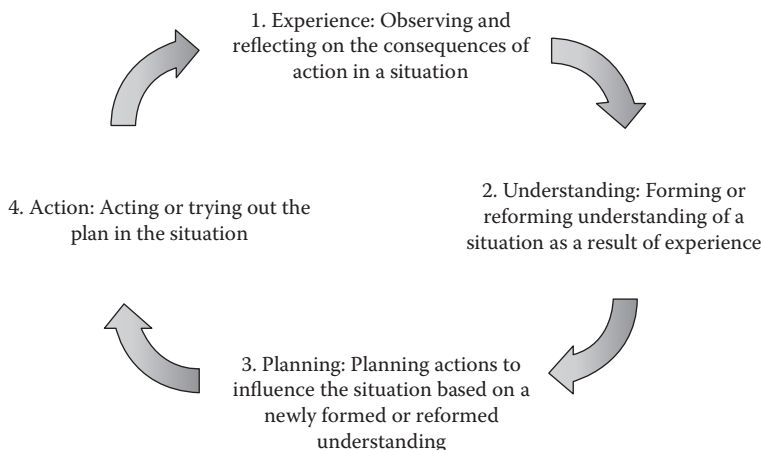
|                                                   |     |
|---------------------------------------------------|-----|
| Embedding E/HF in the Organisational System ..... | 976 |
| Current Challenges for E/HF Professionals.....    | 977 |
| Increasing Complexity in Modern Systems.....      | 978 |
| Trends in E/HF in the Next Decade.....            | 979 |
| Conclusion .....                                  | 981 |
| References.....                                   | 982 |

In the conclusion to the third edition of this book, John Annett (2005) described the evolution of ergonomics as a science from its heritage in military and engineering contexts. He noted that a discipline that was once focused on the development of methods to make quantitative predictions of human capabilities now considered broader and more applied concepts such as teamwork and communication and worked more closely with subject areas such as human-computer interaction. This evolution has continued and is reflected in the new content within this volume, as well as the change in emphasis of many of the chapters included in this text.

Annett likened the role of the current ergonomist to that of a clinician or architect. So how has ergonomics/human factors (E/HF) developed in the past decade, and how does the practice of an E/HF specialist manifest itself? We consider the job of an E/HF specialist to be that of a 'reflective practitioner'. We hint at this in earlier parts of this book, as John Wilson did in the previous volume and in his writing, considering the notion of E/HF as a 'craft' (Wilson, 2000).

The notion of reflective practice has been implemented within educational theory and practice for many years. For example, Brookfield (1995) states that critical reflection helps us to take informed decisions and develop rationale for practice, and Schon (1983) emphasises the need to theorise within reflective practice. This links closely with an 'action research' approach (as discussed in Chapter 28) where, as shown in the example in Figure 38.1 (Pedler et al., 1986), we move through a cycle of experience, understanding, planning and action in continually reflecting on and improving the way in which we complete tasks, whether those tasks are delivering training or teaching, conducting an evaluation or introducing a new way of working.

So how does reflective practice make the application of E/HF more effective? Management has long embraced 'continuous improvement' – moving beyond specifications of individual changes to a continuous monitoring and recording of performance. Several things have aided this change, and we can see the same thing happening in some contexts within E/HF. Increased presence and practicality of collection of large, detailed sets of data in engineering contexts can now be seen in human operation, work and interaction field. This is demonstrated by the increased emphasis in this book on neuroergonomics and physiological monitoring tools – once for use only in a laboratory context, these previously intrusive and expensive technologies are now becoming practical to implement in the field, allowing us to merge the quantitative approaches that have previously been used



**FIGURE 38.1** Action research approach. (From Pedler, M. et al., *A Manager's Guide to Self-Development*, McGraw-Hill, Maidenhead, U.K., 1986.)

for hypothesis-driven and predictive controlled studies, with the emergent hypothesis and systems based, often collaborative and qualitative analysis that can be conducted in the field.

## EMBEDDING E/HF IN THE ORGANISATIONAL SYSTEM

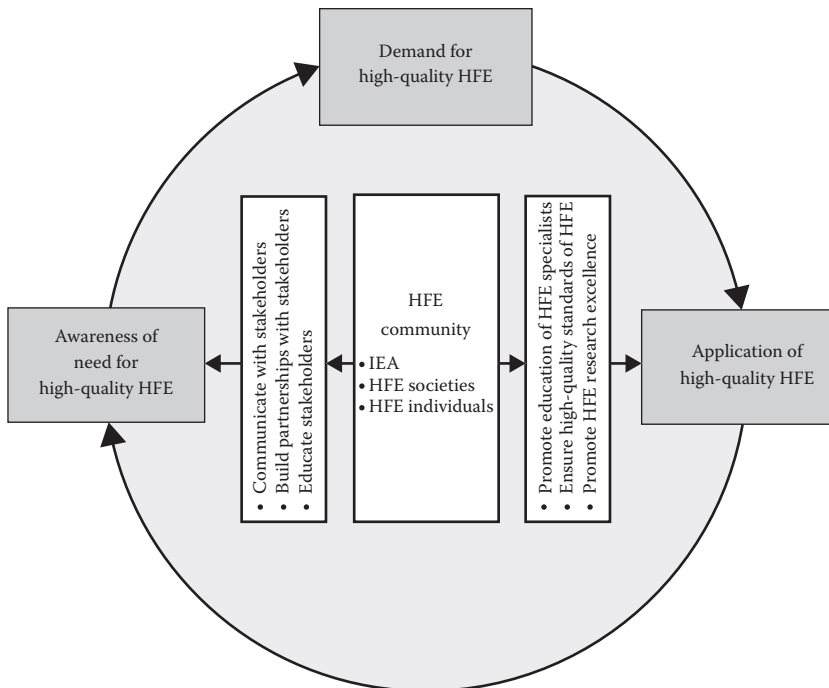
Many of those who are successful at changing the way in which people work, in which products are designed or in which the emphasis of calls for research projects, are focussed are individuals who, whilst experts in E/HF, have also become parts of their organisational structure at a senior level and/or are working with those who are implementing work or work systems on a day-to-day basis.

This approach is particularly encouraged by those who emphasise the performance and quality aims of ergonomics or those that have senior responsibility for the well-being of the workforce. What is important here is that these individuals, even if they are not experts, have a true grasp of the scope of E/HF, its potential for enhancing their organisation and a sustained commitment to ensuring ergonomics becomes part of routine business management. Experience shows that it is often this sustaining or embedding of E/HF that is the hardest challenge. Those organisations that have maintained a long-term commitment are often those who have recognised the need to train representatives of the workforce in basic E/HF thinking, to engage with those who are responsible for performance and quality, and to document and publicise and promote their successes at board level. Without this, there is a real risk that when the E/HF champion or change agent leaves a particular organisation, then the initiative is lost.

In the United Kingdom at least, this situation may be about to change. The recognition of a Royal Charter for the Institute of Ergonomics and Human Factors is imminent as this book goes to press. This recognises that this professional body represents *a field of activity that is unique* and not covered by other professional bodies. We believe that affording this institute such recognition provides a much-needed professional platform on which to build. Larger organisations will hopefully now understand that a subject as complex as ergonomics/human factors cannot be left to those with only a minimal or superficial understanding of the subject or those for whom it is considered as an 'extra' brief to carry. It will be important to assess the longer-term impact of this charter and share the outcomes with the international ergonomics community.

The value of ergonomics for stakeholders has been described in an important recent paper on the future of ergonomics (Dul et al., 2012). This multi-author paper provides both a vision and an approach for the future of ergonomics. Consensus views as to the value of E/HF to systems actors (e.g. the workforce), system products, systems decision makers, system influencers and system





**FIGURE 38.2** Human Factors/Ergonomics demand development cycle. (From Dul, J. et al., *Ergonomics*, 55, 377, 2012.)

experts have been identified and documented. Those embedded in work organisations at a senior level may wish to refresh themselves on these values and ensure that their colleagues appreciate their relevance and benefits. In this way, the demand for high-quality ergonomics is strengthened. The reinforcing relationship between awareness of, demand for and application of high-quality ergonomics demonstrates a way forward that we endorse (Figure 38.2).

Some excellent examples of this approach already exist, from the notion of ergonomists aligned to individual production lines within automotive manufacturing in Ford, to appointments of HF experts within hospital contexts, to dedicated research units aligned with key activities (such as has commonly been the case in the defence sector in particular), from those with ergonomics backgrounds rising to senior management positions in multinational corporations. This works as an approach, but is not particularly ‘scientific’. It is also useful and important to be able to demonstrate the effectiveness and value of such an approach in the representation of outcomes of interventions in a clear, often quantitative form (as discussed in Chapter 35 of this book), and the value and contribution of such an ‘embedded’ approach is often more evident in changes in culture than in specific examples of interventions. Perhaps it is useful to consider E/HF as a ‘continuous state’ rather than a series of ‘discrete events’.

Sometimes we hear the argument from colleagues that ‘human factors is common sense’ (as discussed in Chapter 1). Well, we would argue that the *concept* of E/HF is indeed common sense – of course, it is sensible to implement interventions that lead to a better work design, better work performance and better work culture. But the selection, implementation and interpretation of the methods that we use are *not* common sense. Hence the value of this text.

## CURRENT CHALLENGES FOR E/HF PROFESSIONALS

Much of the early data gathering of human performance and capability was limited to narrowly defined demographic groups (e.g. young male army personnel, college students, laboratory subjects). These data were used to assist the design of work and workplaces and undoubtedly provided

much needed information leading to better workspaces, displays and controls. Sadly, as these data were applied in other contexts and with other demographics, their usefulness diminished. It rapidly became evident that other groups/populations do not necessarily behave in the same way and that the variations between groups and individuals are rarely predictable. Accounting for individual variability became a huge challenge, especially in complex systems that placed competing demands (i.e. physical, psychological and social) on those operating within it.

Attempting to understand this variability in human populations (and the underlying statistically described distributions) leads to the realisation that 'outliers' may exhibit characteristics and behaviours that are far from predictable and that when combining distributions, we may have little to guide us on what to expect. It is here that the reflective practitioner model becomes of great importance.

For example, we need to recognise the value of the unexpected, the sentinel observation that can lead to new ways of understanding a problem or of identifying a potential solution. A recent study of telecare for patients with dementia offers one such example. A family with an ageing mother with dementia had become concerned that she was leaving the house through the front door at night and becoming lost and confused. The usual selection of state of the art and costly technological solutions were advanced. Instead, with the help of the family and a dementia adviser, a different, non-technological solution was found that was based on the understanding of the psychological capabilities of that individual and the residual long-term memory recall patterns that she had. The solution, the hanging of a sign stating 'Gents Toilet' on the front door at night, provided a simple, yet effective, barrier to an unsafe practice for that individual. It also serves to remind ergonomists of the need to understand the complexity of systems and individual differences and that the rush towards technological solutions may be misguided.

The 'craft' of the ergonomist has therefore become one of understanding these challenges, working with the 'hard data' but accepting their limitations yet moving towards improvements of the system with the help of those who have other important sources of knowledge, including those operating within it. This role of facilitating improvements in a reflective and flexible yet pragmatic manner describes the day-to-day life of many experienced ergonomists.

## INCREASING COMPLEXITY IN MODERN SYSTEMS

Understanding complexity in modern systems is a theme of many discussions within this book. Yet most ergonomists still struggle to come to terms with the need to map in some form the complex world they may be investigating or inhabiting. We have long argued that ergonomists must articulate what they have and what they have not examined in the system under investigation. Mapping and visualisation techniques may be helpful in this (perhaps using techniques such as those described in Chapters 10 and 34, on inclusive and participatory design). We have found that mapping systems with end users can provide at times dramatic insights into barriers and 'fault lines' in organisations. Often these manifest themselves as communication barriers that have hitherto been obscured. Many organisations believe they have process maps and systems descriptions that suffice in this regard. However, increasingly complex systems that advance in unpredictable ways often see 'standard operating procedures', and other user guidance fall into disuse. The challenge for future ergonomists is to adapt their practice to these challenges and articulate and present their analysis and solutions in a professionally authoritative manner. We also anticipate the greater use of such techniques in the future for more effective commissioning of services and of performance management in complex sectors. For example, health and social care delivery is dependent on a full and systematic analysis of each member of the system as well as that of the care recipient (i.e. patient).

Much of E/HF work and investment in industrial practice has come about in the aftermath of major incidents. Every E/HF student is aware of the examples of Three Mile Island, Bhopal, Chernobyl, the Challenger Space Disaster, Kegworth, Clapham Junction, Zeebrugge, Überlingen and Ladbroke Grove. Examples of incidents are also engaging and serve to 'sell' the importance of E/HF in domains ranging from health care to nuclear power plant operations. It is a sad fact that

the occurrence of major incidents such as these raises the prominence of the role of human factors, of human error (see Barry Kirwan's Chapter 30 for a useful note of caution on the use of the term 'human error') and often directly lead to significant investment in fundamental research and practitioner roles working directly with industry. This has in many cases led to some excellent and transformative work – for example much of Helen Muir's research on aircraft evacuation emerged from the incident at Manchester airport when a fire occurred, and many passengers failed to safely escape.\*

There are also challenges with using incidents as a major motivator for research, as they can lead to focus on a single element of the system, rather than the whole systems approach. There is therefore an increasing move within the E/HF community to move beyond the analysis of negative incidents – in safety research, this has been termed as moving from 'Safety I' to 'Safety II' (Hollnagel, 2014). This new approach has many positive benefits but presents interesting methodological challenges. The main challenge can be summed up as 'how can we prove the null hypothesis?' – essentially, how can we demonstrate that the lack of a major incident resulted from effective E/HF-led intervention? For example, when Air France Flight 358 crashed just after landing at Toronto Pearson International Airport in 2005, all passengers survived. The incident report particularly noted the good work of the on-aircraft staff who marshalled the passengers to safety, but made no mention of the potentially life-saving role of the aircraft design, lights to guide passengers to exit, design of exit doors and presentation of safety information, all of which have significantly improved from an E/HF perspective since the Manchester airport accident. The field of resilience (Hollnagel et al., 2011) goes some way to developing approaches to capture the elements of a resilient system, but as highlighted in Chapter 35 on cost–benefit analysis, as a discipline we have a difficult task if we wish to demonstrate the role of good design, participation and a systems approach in leading to better performance, better work design and better work culture.

## TRENDS IN E/HF IN THE NEXT DECADE

It is of course a highly risky task to attempt to predict the future,† especially in a text such as this that is normally only revised once a decade or so, but it is useful to consider how changes in society might impact on the practice of E/HF in the short- and medium-term future.

The first trend is associated with automation and the move from automation to autonomy. The 'robot revolution' predicted in the 1970s did not perhaps materialise to the extent that was predicted. In many manufacturing contexts, 'full' automation has indeed been implemented, with the role of the human being as one of a monitor who deals with unexpected or unusual situations, as predicted by Lisanne Bainbridge in her 'Ironies of Automation' (Bainbridge, 1983). However, we are now seeing increased prevalence of autonomous systems, particularly in the transport domain, where driverless trains now exist in a number of contexts, and the first driverless cars are undergoing field trials. This presents new challenges for E/HF evaluation in the need to not only capture the interaction which is likely to even be less visible or articulable than some of the examples described in Chapters 3 and 7 of this book, but will also affect an even wider number of individuals. The notion of 'user' is difficult to define in such contexts – for example – who is the 'user' of a driverless car? The person sitting in the 'driving' seat? The person sitting in the rear of the vehicle? Other road users? Pedestrians? Those monitoring the traffic in the local area? The local council or government? The necessity to take a systems approach becomes clear, as does the need for E/HF methods to retain a socio-technical approach that includes interactions, communications,

---

\* Helen Muir is quoted in the *Times Educational Supplement*, 28 August 1998. 'It's a terrible thing, but the Manchester air disaster made my career.'

† A favourite example of mine (Sarah Sharples) of this comes from a book I had as a child, titled *Living in the Future*. Amongst many accurate predictions about changes in communications and transport systems is the suggestion that 'Hundreds of years ago people worked from dawn until dusk. Now, in Britain for example, people work about 7 or 8 hours a day and this is likely to fall in the future'. So far, this reduction in working hours has failed to materialise.

attitudes and expectations within the elements of interest. The role of the social and societal element of 'fitting the task to the person' has long been acknowledged, but technological evolutions deem it critical.

As noted throughout this book, and in particular in Chapters 21 and 22, which deal with psychophysiological methods and neuroergonomics, physiological sensing is becoming more affordable and less intrusive; it is now practicable to introduce such sensors into workplace context. This presents interesting ethical challenges in terms of the extent to which we should monitor and store data about workplace performance. E/HF practitioners have always had to be open about their collection of data for the purposes of research or design interventions, and normally lay down clear ground rules about how that data will be used in the future both by themselves and, in particular, by any employers of those from whom data have been collected. In the same way as we currently ensure confidentiality and anonymity of data collected in interviews and research studies, we need to put in place measures to ensure that any physiological data collected in a workplace context are stored and used in a way that protects the privacy of the individual. In addition, we need to continue to work to recognise the extent to which such data can be used for the prediction of future performance and individual well-being, and retain established evaluation and monitoring approaches alongside these technology-led methods. Work in computer science (Sweeney, 2000) has demonstrated that individuals can be identified from relatively few data points; the E/HF community needs to be aware of these issues and manage data collection appropriately.

The nature of such sensing methods will also change. Most, if not all, of the monitoring methods presented in this book involve some level of intrusion, whether it is the visible presence of an observer in the work environment or the participant's awareness of additional sensing equipment being worn on their person. Advances in computer vision (Vinciarelli et al., 2009) mean that we are increasingly able to make inferences about activity, interaction and emotion from remotely captured video data. Again, this presents ethical challenges, both in the interpretation and use of these data, and ensuring that individuals are aware of how such data are being used to inform E/HF analysis. This opportunity for additional and large sets of data collection is also accompanied by societal changes. We are increasingly accepting the assumed presence of closed-circuit television in a range of types of public spaces and are beginning to understand how our data are stored in our 'lifelong contextual footprint' as we shop, study or socialise using electronic fora.

These technology-driven methods, along with our continual production of and interaction with information throughout our lives, lead to large sets of what are currently being termed 'big data'. Definitions of big data vary, but generally consider data sets that are not possible to be processed by conventional individual computing tools and therefore require novel, often distributed and algorithm-driven approaches to their filtering and interpretation (see Mayer-Schonberger and Cukier, 2013). The challenges of big data for the E/HF community are still emerging, but it is clear that we must be aware of such developments and are likely to have to particularly review the methods we currently use for statistical analysis, visualisation and interpretation of data when working with big data to inform E/HF design and analysis.

Traditionally, we have talked about ergonomics as 'fitting the task to the person' and generally considered this in a workplace context. Indeed, the Greek derivation of the word ergonomics is from 'ergates', meaning work, and 'nomos', meaning law. The work-life boundary continues to blur, and future E/HF methods may need to be equally applicable to the home context as well as in a workplace setting. This may, for example, demand more of a focus on methods that can be applied in a longitudinal or remote form (such as remote diary studies or automatic data collection techniques) and also presents ethical challenges around the permissions and capture of data from those who are also inhabiting a social or family context of a participant of interest during a study.

The progress of technology has encouraged systems that enable remote working, distributed work systems and global working hours with little allowance made for home/work delineation. The social consequences of this have yet to be fully realised. What is apparent is that the social fabric that bonds a workforce or a work sector is being rediscovered. The design community that has

'sprung up' in the East End of London is reminiscent of other industries (e.g. clothing) that were its forerunners in recognising the synergistic advantages that lay in close co-location. Ergonomists in future are likely to need to pay greater attention to the 'socio' element of a socio-technical system approach and work alongside those with specialist knowledge in this area.

Globalisation of work, afforded through the Internet in particular, represents an opportunity for ergonomists to extend their skill sets yet further. How will participatory design approaches differ in different cultural settings? How might such findings transfer across cultural boundaries? How might the need to work with more frugal resources lead to innovative solutions? How will the voice of the end user be appropriately heard in such global workforces? We believe the professional ergonomist has a vital role to play in meeting such challenges.

## CONCLUSION

It is clear that E/HF is becoming increasingly recognised as an important concept, but perhaps as a community, we are less competent at communicating the value and nature of good E/HF approaches to colleagues from other disciplines.

However, some concepts that have been accepted since the 1950s remain. To succeed in implementing E/HF, we must work *with* our colleagues from other disciplines, by communicating in language that they understand and producing tools that can be used appropriately and effectively by those with different backgrounds and experience. As Hwyl Murrell is often quoted as saying (Waterson and Sell, 2006), what the world needs is ergonomics, not ergonomists. We would argue that we need both; ergonomists are able to communicate concepts and develop new approaches, but perhaps most importantly of all, E/HF practitioners have a particular, and sometimes hard to articulate, way of 'thinking about a problem'. This goes back to the concept of ergonomics as a 'craft', or, to use different terminology, a 'state' rather than a 'series of discrete events'. But we need to marry this creative approach with support from a scientific approach.

It is our belief that one of the things that readers of this book should achieve, thanks to the learned contributions of authors from such diverse backgrounds in research and practice, from different geographical settings and taking different perspectives, is the improved knowledge to help them 'think' about E/HF. This returns to our theme of reflective practice. Many E/HF practitioners (including the authors of this chapter!) freely admit to 'thinking about ergonomics all the time'; we have both been in situations where, in the middle of the conversation about something completely different, we and an E/HF colleague have been completely distracted by the way that we observe someone working or the way in which something is designed. We are already, perhaps without realising it, reflective practitioners.\* This is an approach that has been tacitly assumed for many years amongst the E/HF community, yet we have very few tools to enable us to reflect on our own thoughts and knowledge and use our rich experience from our work and lives to enhance our practice. We could benefit from learning from other professionals, such as in education or counselling, to help us to formalise and enhance our reflective practice approach. Tools are starting to exist and enable us to share our thoughts and ideas in ways that were not previously possible. We can use digital tools to create reflective online diaries and use blogs and social media to communicate and discuss our thoughts in a less formal manner than the traditional journal article or conference submission. As a discipline, we note that we are perhaps slightly more reticent than our colleagues from disciplines such as the social sciences or computing to embrace these tools (perhaps due to the inherently confidential and personal nature of much of our work that can prevent us from sharing details of our work in some circumstances), but we should work to understand the value of such approaches to our continually improving practice. This can not only ensure that those of us in

---

\* It is interesting to think about the evidence for this reflective practice in action. My (Sarah Sharples) personal smartphone, for example, is filled with photos of confusingly designed buttons, taps and doors, doodles on whiteboards and extracts from papers that I've seen.

research and practice settings can enhance our work, but also that we can more explicitly teach our students and junior colleagues the 'art' or even 'passion' of working in E/HF.

This book goes some way to addressing this challenge; by thinking about and understanding the context of application of methods, E/HF professionals should be able to evolve as individuals, reflect on what tools work and do not work, in which circumstances and why. Many of the authors of the chapters in this book have brought exactly this type of experience to their descriptions and analysis of the methods presented. This volume has always aspired not to be a cookbook of ergonomics methods, but to place ergonomics methodologies in context.

By acknowledging the notion of E/HF as reflective practice, we can continue to develop appropriate methods and tools, along with the underlying theoretical frameworks that support the translation of our work to different domains. E/HF may not always be about delivering the 'right' solution(s), but is always about delivering the right approach.

## REFERENCES

- Annett, J. (2005). Conclusions. In J.R. Wilson and E.N. Corlett (eds.), *Evaluation of Human Work*. London, U.K.: Taylor & Francis Group.
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19, 775–779.
- Brookfield, S.D. (1995). *Becoming a Critically Reflective Teacher*. San Francisco, CA: Jossey-Bass Ltd.
- Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W.S. and van der Doelen, B. (2012). A strategy for human factors/ergonomics: Developing the discipline and profession. *Ergonomics*, 55, 377–395.
- Hollnagel, E. (2014). *Safety-I and Safety-II: The Past and Future of Safety Management*. Farnham, U.K.: Ashgate.
- Hollnagel, E., Pariès, J., Woods, D.D. and Wreathall, J. (eds.) (2011). *Resilience Engineering Perspectives*. Vol. 3: *Resilience Engineering in Practice*. Farnham, U.K.: Ashgate.
- Mayer-Schonberger, V. and Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton, Mifflin, Harcourt.
- Pedler, M., Burgoyne, J. and Boydell, T. (1986). *A Manager's Guide to Self-Development*. Maidenhead, U.K.: McGraw-Hill.
- Schon, D.A. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Sweeney, L. (2000). Simple demographics often identify people uniquely (Data Privacy Working Paper 3). Pittsburgh, PA: Carnegie Mellon University.
- Vinciarelli, A., Pantic, M. and Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12), 1743–1759.
- Waterson, P. and Sell, R. (2006). Recurrent themes and developments in the history of the Ergonomics Society. *Ergonomics*, 49(8), 743–799.
- Wilson, J.R. (2000). Fundamentals of ergonomics. *Applied Ergonomics*, 31, 557–567.



# EVALUATION OF HUMAN WORK

## FOURTH EDITION

Edited by  
**JOHN R. WILSON and SARAH SHARPLES**

**“From the first edition, this text was a ‘must have’; it still is. ... The emphasis on being up to date and authoritative makes the book an essential source for anyone interested in the human–system design and interaction.”**

—Professor Ann Williamson, University of New South Wales, Australia

Written by experts with real-world experience in applying ergonomics methodology in a range of contexts, *Evaluation of Human Work, Fourth Edition* explores ergonomics and human factors from a ‘doing it’ perspective. More than a cookbook of ergonomics methods, the book encourages students to think about which methods they should apply, when and why. Now in its fourth edition, the editors have again brought together a team of world-renowned experts and created a forum for them to discuss and debate their most valued approaches.

### SEE WHAT’S NEW IN THE FOURTH EDITION:

- Significant change of authorship and content in each chapter
- Five new chapters covering inclusive design, situation awareness, neuroergonomics, ergonomics and quality and standards in ergonomics design and evaluation
- Discussion of the idea of reflective practice and Ergonomics/Human Factors
- Examination of the systems perspective now prevalent in the field
- Coverage of new tools and the relationship between these physiologically measured parameters and theoretical concepts such as workload, fatigue, stress or effort

In addition to being a foundational text, the book is an authoritative reference for practising ergonomists. Completely revised and updated, this fourth edition retains the features that have made the book a perennial favorite. After reading each chapter, you will feel as though you understand the context of how to practise ergonomics and what aspects of method selection and implementation you should consider. The book provides a guide to the benefits ergonomics and human factors can offer industry, business or human service professionals.

WITH VITALSOURCE®  
**EBOOK**



- Access online or download to your smartphone, tablet or PC/Mac
- Search the full text of this and other titles you own
- Make and share notes and highlights
- Copy and paste text and figures for use in your own documents
- Customize your view by changing font size and layout



**CRC Press**  
Taylor & Francis Group  
an **informa** business

[www.taylorandfrancisgroup.com](http://www.taylorandfrancisgroup.com)

6000 Broken Sound Parkway, NW  
Suite 300, Boca Raton, FL 33487  
711 Third Avenue  
New York, NY 10017  
2 Park Square, Milton Park  
Abingdon, Oxon OX14 4RN, UK

**K15856**

ISBN: 978-1-4665-5961-5



[www.crcpress.com](http://www.crcpress.com)

